

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:31:16 ; Search time 34 Seconds
(without alignments)
1478.747 Million cell updates/sec

Title: US-09-584-411c-22
Perfect score: 2017
Sequence: 1 MLHAANKGKPSAEAGRP.....NDGDKEMVSFTVVLDTGI 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA.*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2017	100.0	381	10	US-09-808-602-6
2	2017	100.0	381	11	US-09-808-602-6
3	2017	100.0	381	11	US-09-808-602-6
4	2017	100.0	381	11	US-09-808-602-6
5	2017	100.0	381	12	US-09-808-602-6
6	1931	95.7	2724	11	US-09-808-602-13
7	1931	95.7	2724	11	US-09-808-602-13
8	1931	95.7	2724	11	US-09-808-602-8
9	1931	95.7	2724	11	US-09-808-602-8
10	1889.5	93.7	2764	11	US-09-808-602-80
11	1889.5	93.7	2764	11	US-09-808-602-80
12	1889.5	93.7	2765	11	US-09-808-602-84
13	1889.5	93.7	2765	11	US-09-808-602-84
14	1820.5	90.3	2802	11	US-09-808-602-81
15	1820.5	90.3	2802	11	US-09-808-602-81

16	793.5	39.3	2771	10	US-09-808-602-82	Sequence 82, Appl
17	793.5	39.3	2771	11	US-09-808-602-82	Sequence 70, Appl
18	790.5	39.2	560	15	US-10-290-578-4	Sequence 4, Appl
19	790.5	39.2	768	9	US-09-773-517-11	Sequence 11, Appl
20	790.5	39.2	768	9	US-09-773-517-11	Sequence 11, Appl
21	790.5	39.2	768	9	US-09-849-868-11	Sequence 11, Appl
22	790.5	39.2	768	10	US-09-808-602-85	Sequence 85, Appl
23	790.5	39.2	768	15	US-10-290-578-2	Sequence 2, Appl
24	658	32.6	501	15	US-10-290-578-10	Sequence 10, Appl
25	129.5	6.4	731	14	US-10-086-464-17	Sequence 17, Appl
26	124.5	6.2	731	14	US-10-086-464-8	Sequence 8, Appl
27	124.5	6.2	4019	10	US-09-738-973-425	Sequence 425, App
28	124.5	6.2	4019	10	US-09-854-133-425	Sequence 425, App
29	124.5	6.2	4019	15	US-10-144-649A-425	Sequence 425, App
30	122.5	6.1	279	14	US-10-001-857-159	Sequence 159, App
31	121	6.0	5179	9	US-09-922-217-1068	Sequence 1068, Ap
32	121	6.0	5179	10	US-09-833-263-1068	Sequence 1068, Ap
33	121	6.0	5179	14	US-10-025-380-1068	Sequence 1068, Ap
34	119	5.9	196	10	US-09-989-920-224	Sequence 224, App
35	118.5	5.9	881	9	US-09-816-860A-2	Sequence 2, Appl
36	117.5	5.8	884	15	US-10-303-683-21	Sequence 21, Appl
37	117.5	5.8	894	11	US-09-291-417-18	Sequence 18, Appl
38	117.5	5.8	894	15	US-10-303-683-20	Sequence 20, Appl
39	117.5	5.8	1367	10	US-09-801-368-108	Sequence 108, App
40	117	5.8	773	15	US-10-225-630-6	Sequence 6, Appl
41	117	5.8	890	15	US-10-158-684-4	Sequence 4, Appl
42	117	5.8	890	15	US-10-158-711-4	Sequence 4, Appl
43	116	5.8	2971	15	US-10-146-473-50	Sequence 50, Appl
44	115	5.7	495	15	US-10-198-070-48	Sequence 48, Appl
45	115	5.7	833	15	US-10-303-683-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-808-602-6
; Sequence 6, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-6

Query Match 100.0%; Score 2017; DB 10; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHAANKGKPSAEAGRPPTSSPSLLPSAOLPSSHNPVPSVCOMPLDSTSHQIMDT 60
|||||

Db 1 MLHAANKGKPSAEAGRPPTSSPSLLPSAOLPSSHNPVPSVCOMPLDSTSHQIMDT 60
|||||

QY 61 NPDEEFSPNSYLLRACSGPQASSSGPPNNHHSQSTLRPLPPPHNHTLSHHSSANSLSNR 120
|||||

```
Db      61  NPDEEFPNLYLRACGPGQASSGPPNHHSSQTLRPPLPPPHNHTLSHHSSANSNR 120
Qy      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Db      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Qy      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Qy      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Db      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Qy      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
Db      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
Qy      361  YNDGKDKEMVSFNTVVDGTI 381
Db      361  YNDGKDKEMVSFNTVVDGTI 381
```

RESULT 2

```
US-09-977-418-22
; Sequence 22, Application US/09977418
; Publication No. US20030027158A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets et al
; TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded th
; FILE REFERENCE: 15966-552
; CURRENT APPLICATION NUMBER: US/09/977,418
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: USSN 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: USSN 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: USSN 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: USSN 60/201,388
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-418-22
```

```
Query Match      100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDT 60
Db      1  MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDT 60
Qy      61  NPDEEFPNLYLRACGPGQASSGPPNHHSSQTLRPPLPPPHNHTLSHHSSANSNR 120
Db      61  NPDEEFPNLYLRACGPGQASSGPPNHHSSQTLRPPLPPPHNHTLSHHSSANSNR 120
Qy      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Db      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Qy      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Qy      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Db      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
```

```
Db      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Qy      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
Db      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
Qy      361  YNDGKDKEMVSFNTVVDGTI 381
Db      361  YNDGKDKEMVSFNTVVDGTI 381
```

RESULT 3

```
US-09-977-033A-22
; Sequence 22, Application US/09977033A
; Publication No. US20030082554A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Fernandes, Elma
; APPLICANT: Herriman, John
; TITLE OF INVENTION: No. US20030082554A1 polynucleotides and polypeptides encoding human KI
; TITLE OF INVENTION: protein-like and human protein PRO228-like
; FILE REFERENCE: 15966-552 CON-S24
; CURRENT APPLICATION NUMBER: US/09/977,033A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,322
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-033A-22
```

```
Query Match      100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDT 60
Db      1  MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDT 60
Qy      61  NPDEEFPNLYLRACGPGQASSGPPNHHSSQTLRPPLPPPHNHTLSHHSSANSNR 120
Db      61  NPDEEFPNLYLRACGPGQASSGPPNHHSSQTLRPPLPPPHNHTLSHHSSANSNR 120
Qy      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Db      121  NSLTNRRSQIHAPAPNDLATTPEVSQVLDQSVLNSNVPLETRHFLFKTSSTPLFSS 180
Qy      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db      181  SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Qy      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Db      241  YFTVPMSLKNSSIDSGEAEVGRVTOEVPVFWRSQIHLISQPFKFNISLGDALFGV 300
Qy      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
Db      301  YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQVLDVGLWHLAF 360
```

Db 301 YIRRLPSPAQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
QY 361 YNDGDKEMVSNVTYVLDGTI 381
|||||
Db 361 YNDGDKEMVSNVTYVLDGTI 381

RESULT 4

US-09-800-198-6
; Sequence 6, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-6

Query Match 100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSHNPVPVSCOMPLDSDNTSHQIMDT 60
|||||
Db 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSHNPVPVSCOMPLDSDNTSHQIMDT 60
|||||
QY 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
Db 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
QY 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLFKTSSTPLFSS 180
|||||
Db 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLFKTSSTPLFSS 180
|||||
QY 181 SSPGYPLTSGTYVTPPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
Db 181 SSPGYPLTSGTYVTPPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
QY 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
Db 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
QY 301 YIRRLPSPAQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
Db 301 YIRRLPSPAQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
QY 361 YNDGDKEMVSNVTYVLDGTI 381
|||||
Db 361 YNDGDKEMVSNVTYVLDGTI 381

RESULT 5

US-09-977-751C-22
; Sequence 22, Application US/09977751C
; Publication No. US20030134430A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A

; APPLICANT: Fernandes, Elma
; APPLICANT: Herrman, John
; APPLICANT: Vernet, Corine
; TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhab
; TITLE OF INVENTION: Polypeptides.
; FILE REFERENCE: 15966-552 CON S-40
; CURRENT APPLICATION NUMBER: US/09/977,751C
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,322
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-751C-22

Query Match 100.0%; Score 2017; DB 12; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSHNPVPVSCOMPLDSDNTSHQIMDT 60
|||||
Db 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSHNPVPVSCOMPLDSDNTSHQIMDT 60
|||||
QY 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
Db 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
QY 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLFKTSSTPLFSS 180
|||||
Db 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLFKTSSTPLFSS 180
|||||
QY 181 SSPGYPLTSGTYVTPPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
Db 181 SSPGYPLTSGTYVTPPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
QY 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
Db 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
QY 301 YIRRLPSPAQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
Db 301 YIRRLPSPAQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
QY 361 YNDGDKEMVSNVTYVLDGTI 381
|||||
Db 361 YNDGDKEMVSNVTYVLDGTI 381

RESULT 6

US-09-808-602-13
; Sequence 13, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S

QY 70 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 249
DB 342 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 401
QY 250 NSSIDSGAEVGRRTVQEVPPGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPS 309
DB 402 NSSIDSGAEVGRRTVQEVPPGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPS 461
QY 310 HAQYDFMERLDGKEKWSVVEPRRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 369
DB 462 HAQYDFMERLDGKEKWSVVEPRRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 521
QY 370 VSFNTVWLD 378
DB 522 VSFNTVWLD 530

RESULT 9

US-09-800-198-8
; Sequence 8, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-8

Query Match 95.7%; Score 1931; DB 11; Length 2733;
Best Local Similarity 98.9%; Pred. No. 7.4e-143;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 10 KPSAAGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 69
DB 162 KSDDNGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 221
QY 70 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 249
DB 342 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 401

QY 250 NSSIDSGAEVGRRTVQEVPPGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPS 309
DB 402 NSSIDSGAEVGRRTVQEVPPGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPS 461
QY 310 HAQYDFMERLDGKEKWSVVEPRRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 369
DB 462 HAQYDFMERLDGKEKWSVVEPRRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 521
QY 370 VSFNTVWLD 378
DB 522 VSFNTVWLD 530
RESULT 10
US-09-808-602-80
; Sequence 80, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 2764
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-808-602-80

Query Match 93.7%; Score 1889.5; DB 10; Length 2764;
Best Local Similarity 88.5%; Pred. No. 1.4e-139;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;
QY 10 KPSAAGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 69
DB 162 KSDDNGRPIPTSSSULLPSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 221
QY 70 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSQTLPRLPPPPHNTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFI----- 243
DB 342 GTVYTPPRLPRNTFSRKAFKKPKSKYCSWKCAALSAIAAALLLAILLAYFIAMHLG 401
QY 244 -----VPMKSNSSIDSGAEVGRRTVQEV 268
DB 402 LNWQLQPADGHTFNGVRTGLPGNDVATVPSGKVPWSLKNSSIDSGAEVGRRTVQEV 461
QY 269 PGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPSHQAQYDFMERLDGKEKWSV 328
DB 462 PGVFWRSQIHISQOFLKFNISLQKDALFGVYIRRLGPPSHQAQYDFMERLDGKEKWSV 521
QY 329 ESPRERRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKEMVSFNTVWLD 378
DB 522 ESPRERRSTQTLVQNEAVFQYLDVGLHFLAFYNDGKEMVSFNTVWLD 571

RESULT 11

US-09-800-198-68
; Sequence 68, Application US/09800198
; Publication No. US20030087816A1

GENERAL INFORMATION:

; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca

; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

; FILE REFERENCE: 15966-697

; CURRENT APPLICATION NUMBER: US/09/800,198

; CURRENT FILING DATE: 2001-03-05

; PRIOR APPLICATION NUMBER: 60/186,596

; PRIOR FILING DATE: 2000-03-03

; NUMBER OF SEQ ID NOS: 98

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 68

; LENGTH: 2764

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-800-198-68

Query Match 93.7%; Score 1889.5; DB 11; Length 2764;

Best Local Similarity 88.5%; Pred. No. 1.4e-139;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

```
QY 10 KPSAAGRPPTPTSSSLPSAQLPSSHNPPVSCOMPLDSTSHQIMDTPDEEFSN 69
      | : |||||
Db 162 KSDDNGRPPTPTSSSLPSAQLPSSHNPPVSCOMPLDSTSHQIMDTPDEEFSN 221
      | : |||||
QY 70 SYLLRACSGPQQAASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 129
      | : |||||
Db 222 SYLLRACSGPQQAASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 281
      | : |||||
QY 130 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLKTSSTGSPPLFSSSPGYPLTS 189
      | : |||||
Db 282 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLKTSSTGSPPLFSSSPGYPLTS 341
      | : |||||
QY 190 GTVYTPPRLPNTFSRKAFKLKPKSKYCSWKAALSAIAAALLAILLAYFI----- 243
      | : |||||
Db 342 GTVYTPPRLPNTFSRKAFKLKPKSKYCSWKAALSAIAAALLAILLAYFIAMHLG 401
      | : |||||
QY 244 -----VPWSLKNSIDSIDSGEAEVGRRTQEV 268
      | : |||||
Db 402 LNWQLPADGHTFNNGVRTGLPGNDVDVATVPSSGKVPWSLKNSIDSIDSGEAEVGRRTQEV 461
      | : |||||
QY 269 PGVFWRSQIHISQOFLAFNLSIGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 328
      | : |||||
Db 462 PGVFWRSQIHISQOFLAFNLSIGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 521
      | : |||||
QY 329 ESPRERSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 378
      | : |||||
Db 522 ESPRERSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 571
      | : |||||
```

RESULT 12

US-09-808-602-84

; Sequence 84, Application US/09808602

; Patent No. US20020155115A1

GENERAL INFORMATION:

; APPLICANT: Vernet, Cornie A

; APPLICANT: Fernandes, Elma

; APPLICANT: Shimkets, Richard A

; APPLICANT: Herrman, John L

; APPLICANT: Majumder, Kumud

; APPLICANT: Mishra, Vishnu

```

; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-800-198-72

Query Match
Best Local Similarity 93.7%; Score 1889.5; DB 11; Length 2765;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAEAGRIPTSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSN 69
DB 162 KSDDDNGRPIPTSSSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSN 221
QY 70 SYLLRACSGPOAASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQ 129
DB 222 SYLLRACSGPOAASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQ 281
QY 130 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 341
QY 190 GTVYTPPPRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 243
DB 342 GTVYTPPPRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 401
QY 244 -----VPWSLNKSSIDSGEAEVGRRTQEV 268
DB 402 LNWQLQPADGHTFNGVRTGLPGNDVATVPSGKVPWSLNKSSIDSGEAEVGRRTQEV 461
QY 269 PPGVFWRSQIHSIQPOFLKFNISLGKDALFGVYIRRGLPSPSHQAQYDFMERLDGKESWV 328
DB 462 PPGVFWRSQIHSIQPOFLKFNISLGKDALFGVYIRRGLPSPSHQAQYDFMERLDGKESWV 521
QY 329 ESPRRRSQIOTLVQNEAVFVQYLDVGLWHLAFYNDGDKKEMVSPNTVVD 378
DB 522 ESPRRRSQIOTLVQNEAVFVQYLDVGLWHLAFYNDGDKKEMVSPNTVVD 571

RESULT 14
US-09-808-602-81
; Sequence 81, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NO. US20020155115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 2802
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-808-602-81

Query Match
Best Local Similarity 90.3%; Score 1820.5; DB 10; Length 2802;
Matches 346; Conservative 9; Mismatches 7; Indels 41; Gaps 1;

QY 17 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSNVLRLAC 76
DB 197 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSNVLRLAC 256
QY 77 SGPOQASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQIHAPAPA 136
DB 257 SGPOQASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQIHAPAPA 316
QY 137 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 196
DB 317 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 376
QY 197 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 244
DB 377 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 436

Query Match
Best Local Similarity 85.9%; Score 1820.5; DB 11; Length 2802;
Matches 346; Conservative 9; Mismatches 7; Indels 41; Gaps 1;

QY 17 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSNVLRLAC 76
DB 197 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPNDEEFSNVLRLAC 256
QY 77 SGPOQASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQIHAPAPA 136
DB 257 SGPOQASSGPPNHSQSTLRPPPPHNHTLSHHSSANSLSNRSQIHAPAPA 316
QY 137 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 196
DB 317 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 376
QY 197 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 244
DB 377 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 436
```

QY	245	-----PWSLKNSIDSGEAEVGRRTQEVPPGVFWR	275
Db	437	ADGHTESNGLRPGAAEDGAAAPAGRGFWTRNNSIDSGETEVGRKVTQEVPPGVFWR	496
QY	276	SOIHISQOFLKFNISLKGKDALFGVYIRRGLPSPSHAQYDFMERLDGKEKWSVYESPRRR	335
Db	497	SOIHISQOFLKFNISLKGKDALFGVYIRRGLPSPSHAQYDFMERLDGKEKWSVYESPRRR	556
QY	336	SIQTLVQNEAVFVOYLDVGLWHLAFYNDGKDKEMVSFNTVWLD	378
Db	557	SIQTLVQNEAVFVOYLDVGLWHLAFYNDGKDKEMVSFNTVWLD	599

Search completed: August 26, 2003, 14:33:31
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:27:38 ; Search time 44 Seconds
(without alignments)

1374.428 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHAANKGRKPSAAGRPIP.....NDGKDKEMVSFNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2017	100.0	381	22	AAU08679
2	2017	100.0	381	22	AA861141
3	2005	99.4	429	22	AAU00392
4	1931	95.7	2724	22	AAU08681
5	1931	95.7	2733	22	AAU08680
6	992.5	49.2	865	22	ABG03234
7	976.5	48.4	2721	23	ABP53587
8	972.5	48.2	2613	23	ABP53589
9	972.5	48.2	2628	23	ABP53588
					Human FCTR3a polyp
					Human NOV11 prote
					Human secreted pro
					Human FCTR3f polyp
					Human FCTR3b polyp
					Novel human diagno
					Human NOV15b prote
					Human NOV15d prote
					Human NOV15c prote

10	956.5	47.4	2725	23	ABP53586	Human NOV15a prote
11	793	39.3	2758	23	ABG97359	Human CGDD10, INCY
12	792.5	39.3	2769	23	ABG70388	Human TEN-M4-like
13	790.5	39.2	768	19	AAW44817	Human gamma-heregu
14	790.5	39.2	768	20	AAW06639	Gamma-herequin.
15	790.5	39.2	768	21	AAW71176	Human Heregulin va
16	790.5	39.2	768	22	AAU09891	Human Heregulin, g
17	783	38.8	2794	23	ABN8401	Human NOV1, a TEN-
18	747	37.0	2136	22	AAW78695	Human protein SEQ
19	739	36.6	2725	23	ABG61913	Prostate cancer-as
20	665	33.0	527	19	AAW4818	Human gamma-heregu
21	429.5	21.3	260	22	ABG05089	Novel human diagno
22	241.5	12.0	97	22	ABG03232	Novel human diagno
23	186.5	9.2	2515	22	ABW71354	Drosophila melanog
24	150	7.4	433	23	ABP69330	Human polypeptide
25	147	7.3	477	23	ABP41567	Human ovarian anti
26	144	7.1	777	22	ABG65598	Drosophila melanog
27	137.5	6.8	598	22	ABG14000	Novel human diagno
28	136.5	6.8	1895	22	ABW70088	Drosophila melanog
29	136	6.7	459	23	ABP43882	Human EXMAD-20 pro
30	136	6.7	571	22	AAW27242	Human EXMAD-20 SEQ
31	134.5	6.7	277	21	AAW42177	Human ORFX ORF1941
32	134.5	6.7	633	22	AAW52321	Las17 protein. Sa
33	134.5	6.7	633	22	AAW67365	Amino acid sequenc
34	134	6.6	1822	23	ABP43899	Nuclear protein za
35	132.5	6.6	2703	22	ABG63299	Drosophila melanog
36	132.5	6.6	2703	23	ABG70019	Larval viability a
37	130	6.4	169	22	AAU01990	Gene #26 human sec
38	130	6.4	577	24	ABU00115	Human novel polype
39	129.5	6.4	731	22	AAW74209	Protein encoded by
40	129.5	6.4	731	23	ABW93202	Herbicidally activ
41	129	6.4	237	22	AAW74609	Leishmania major p
42	128.5	6.4	517	23	ABP74037	Candida albicans e
43	128.5	6.4	3232	22	ABW69873	Drosophila melanog
44	128	6.3	356	23	ABW91664	Herbicidally activ
45	128	6.3	516	21	AAW36365	Human TRAF6 bindin

ALIGNMENTS

RESULT 1

AAU08679
ID AAU08679 standard; Protein; 381 AA.

XX AC AAU08679;

XX DT 18-DEC-2001 (first entry)

XX DE Human FCTR3a polypeptide sequence.

XX KW Human; FCTRX; myelogenous leukaemia; carcinoma; melanoma; glioma; astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection; neurological disorder; neurodegenerative disorders; nerve trauma; familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy; demyelinating Gardner syndrome; familial myelodysplastic syndrome; mental health condition; immunological disorder; allergy; infertility; bronchial asthma; Avellino type eosinophilia; lung disease; deafness; reproductive disorder; reproductive disorder; glycoprotein Ia deficiency; desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide; gastric disorders; pancreatic disease; Schistosoma mansoni infection; Spino cerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes; Corneal dystrophy-Greonow type I; Corneal dystrophy-lattice type I; Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive; anti-allergic; antiasthmatic; antifertility; antinflammatory; antidiabetic; prozoocide; hepatotropic; virucide; ophthalmological; KW gynaecological; antinfertility; immunostimulant; auditory; haemostatic; gene therapy; FCTR3a; neurastin-like protein.

XX OS Homo sapiens.

XX PN WO200166747-A2.

XX

PD 13-SEP-2001.
 XX 05-MAR-2001; 2001WO-US07160.
 PF 03-MAR-2000; 2000US-186592P.
 XX 03-MAR-2000; 2000US-186718P.
 PR 06-MAR-2000; 2000US-187293P.
 PR 06-MAR-2000; 2000US-187294P.
 PR 17-MAR-2000; 2000US-190400P.
 PR 07-APR-2000; 2000US-196018P.
 PR 03-JAN-2001; 2001US-259548P.
 XX (CURA-) CURAGEN CORP.
 PA Vernet CAM, Fernandes E, Shinkets RA, Herrmann JL, Majumder K;
 PI Macdougall J, Mishra V, Mezes PS, Rastelli L;
 XX WPI: 2001-596837/57.
 DR N-PSDB; AAS14084.
 XX Novel polypeptides designated as FCTRX polypeptides, useful in
 PT detection, prevention and treatment of a broad range of pathological
 PT states
 XX
 PS Claim 1; Page 33; 215pp; English.
 XX
 CC The invention relates to human FCTRX polypeptides, FCTR1-FCTR7, and the
 CC nucleic acids encoding them. These sequences are useful for the treatment
 CC or prevention of numerous disorders including myelogenous leukaemia,
 CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
 CC allolimmune thrombocytopenia, neurological disorders, neurodegenerative
 CC disorders, nerve trauma, familial myelodysplastic syndrome,
 CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
 CC myelodysplastic syndrome, mental health conditions, immunological
 CC disorders, allergy and infection, bronchial asthma, Avellino type
 CC eosinophilia, lung diseases, reproductive disorders, infertility, male
 CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
 CC desmoid disease, turtot syndrome, liver cirrhosis, hepatitis C, gastric
 CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
 CC infection Spinocerebellar ataxia, Plasmodium falciparum parasitaemia,
 CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
 CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3a, a
 CC neurestin-like protein.
 XX
 SQ Sequence 381 AA;
 Query Match 100.0%; Score 2017; DB 22; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1e-159;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHAANKGRKPSAEAGRIPTTSSPSPSLPSAQLPSSHNPPVPVSCQPLDLSNTSHQIMDT 60
 Db 1 MLHAANKGRKPSAEAGRIPTTSSPSPSLPSAQLPSSHNPPVPVSCQPLDLSNTSHQIMDT 60
 QY 61 NPDEFSNSYLLRACSGPQOASSGPPNHHSSQSLRPLPPPHNHTLSHHSSANSNLR 120
 Db 61 NPDEFSNSYLLRACSGPQOASSGPPNHHSSQSLRPLPPPHNHTLSHHSSANSNLR 120
 QY 121 NSLTNRRSQIHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRHFLKTSSTGSPPLFSS 180
 Db 121 NSLTNRRSQIHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRHFLKTSSTGSPPLFSS 180
 QY 181 SSPGYPLTSGVYTPPPRLLPNTFSRAFKLKPCKSKYCSWKAALSAIAAALLAILLA 240
 Db 181 SSPGYPLTSGVYTPPPRLLPNTFSRAFKLKPCKSKYCSWKAALSAIAAALLAILLA 240
 QY 241 YFIVPWSLKNSSIDSGEAEVGRVTOEYPPGVFWRSQIHISQOPLKFNISLGKDALFGV 300
 Db 241 YFIVPWSLKNSSIDSGEAEVGRVTOEYPPGVFWRSQIHISQOPLKFNISLGKDALFGV 300
 QY 301 YIRGLPSPHAQYDFMERLDGKWKSVVSPRRRSIOTLVQNEAVFVQYLDVGLWHLAF 360
 Db 301 YIRGLPSPHAQYDFMERLDGKWKSVVSPRRRSIOTLVQNEAVFVQYLDVGLWHLAF 360

QY 361 YNDGKDKEMVSFNTVVDGTI 381
 Db 361 YNDGKDKEMVSFNTVVDGTI 381
 RESULT 2
 AAB61141
 ID AAB61141 standard; Protein; 381 AA.
 XX
 AC AAB61141;
 XX
 DT 30-MAR-2001 (first entry)
 XX
 DE Human NOV11 protein.
 XX
 KW Human; NOVX; antiinflammatory; cytostatic; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnarary; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200075321-A2.
 XX
 PD 14-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15303.
 XX
 PR 03-JUN-1999; 99US-0137322.
 PR 16-MAR-2000; 2000US-0189810.
 PR 22-MAR-2000; 2000US-0191158.
 PR 30-MAR-2000; 2000US-0193086.
 PR 31-MAY-2000; 2000US-0137322.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI: 2001-102403/11.
 DR N-PSDB; AAF27859.
 XX
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing
 XX
 PS Claim 1; Page 42-44; 194pp; English.
 XX
 CC The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).
 XX
 SQ Sequence 381 AA;
 Query Match 100.0%; Score 2017; DB 22; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1e-159;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHAANKGRKPSAEAGRIPTTSSPSPSLPSAQLPSSHNPPVPVSCQPLDLSNTSHQIMDT 60

|||||
Db 1 MLHANKRKPSAEAGRIPTSSPSLLPSAQLPSSHNPVPVSCOMPLDNTSHQIMDT 60
QY 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLNNRRSQIHAPAPAPNDLATTPEVSQLOQSWLNSVPLETRHFLFKTSGGSTPLFSS 180
Db 121 NSLNNRRSQIHAPAPAPNDLATTPEVSQLOQSWLNSVPLETRHFLFKTSGGSTPLFSS 180
QY 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCKSWKCAALSAAIAALLAILLA 240
Db 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCKSWKCAALSAAIAALLAILLA 240
QY 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLGKDALFGV 300
Db 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLGKDALFGV 300
QY 301 YIRGLPSPHAQYDFMERLDGKEKWSVVEPRRISQIOTLVQNEAVFYQYLDVGLWHLAF 360
Db 301 YIRGLPSPHAQYDFMERLDGKEKWSVVEPRRISQIOTLVQNEAVFYQYLDVGLWHLAF 360
QY 361 YNDGDKEMVSFNTVWLDGTI 381
Db 361 YNDGDKEMVSFNTVWLDGTI 381

RESULT 3
AAU00392
ID AAU00392 standard; Protein; 429 AA.
XX
AC AAU00392;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human secreted protein, POLY4.
XX
KW Human secreted protein; therapeutic; diagnostic; human; cancer.
XX
OS Homo sapiens.
XX
PN WO200119856-A2.
XX
PD 22-MAR-2001.
XX
PF 13-SEP-2000; 2000WO-US25106.
XX
PR 13-SEP-1999; 99US-0153629.
PR 16-SEP-1999; 99US-0154520.
PR 20-SEP-1999; 99US-0154762.
PR 13-OCT-1999; 99US-0159231.
PR 12-SEP-2000; 2000US-0659634.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Fernandes E, Herrmann JL, Liu X, Yang M, Boldog FL;
XX
DR WPI; 2001-244781/25.
DR N-PSDB; ANS01213.
XX
XX
PT New POLYX polypeptide useful for treating or preventing a POLYX
associated disorder, e.g. cancer -
XX
PS Claim 9; Page 11-13; 152pp; English.
XX

The sequence represents the amino acid sequence of human secreted protein, POLY4. POLYX nucleic acids, polypeptides and antibodies to POLYX can be used for treating or preventing a POLYX associated disorder in a subject, preferably a human. These can be used in the manufacture of a medicament for treating a syndrome associated with a human disease selected from a POLYX-associated disorder, where the therapeutic is a POLYX polypeptide, a POLYX nucleotide or a POLYX antibody. They may also

CC be used to screen for a modulator of activity, or latency, or
CC predisposition to a POLYX associated disorder, e.g. cancer.
XX
SQ Sequence 429 AA;
Query Match 99.4%; Score 2005; DB 22; Length 429;
Best Local Similarity 99.2%; Pred. No. 1.3e-158;
Matches 378; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLHANKRKPSAEAGRIPTSSPSLLPSAQLPSSHNPVPVSCOMPLDNTSHQIMDT 60
Db 1 MLHANKRKPSAEAGRIPTSSPSLLPSAQLPSSHNPVPVSCOMPLDNTSHQIMDT 60
QY 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLNNRRSQIHAPAPAPNDLATTPEVSQLOQSWLNSVPLETRHFLFKTSGGSTPLFSS 180
Db 121 NSLNNRRSQIHAPAPAPNDLATTPEVSQLOQSWLNSVPLETRHFLFKTSGGSTPLFSS 180
QY 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCKSWKCAALSAAIAALLAILLA 240
Db 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCKSWKCAALSAAIAALLAILLA 240
QY 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLGKDALFGV 300
Db 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLGKDALFGV 300
QY 301 YIRGLPSPHAQYDFMERLDGKEKWSVVEPRRISQIOTLVQNEAVFYQYLDVGLWHLAF 360
Db 301 YIRGLPSPHAQYDFMERLDGKEKWSVVEPRRISQIOTLVQNEAVFYQYLDVGLWHLAF 360
QY 361 YNDGDKEMVSFNTVWLDGTI 381
Db 361 YNDGDKEMVSFNTVWLDGTI 381

RESULT 4
AAU08681
ID AAU08681 standard; Protein; 2724 AA.
XX
AC AAU08681;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human FCTR3f polypeptide sequence.
XX
KW Human; FCTR3; myelogenous leukaemia; carcinoma; melanoma; glioma;
astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;
neurological disorder; neurodegenerative disorders; nerve trauma;
familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
demyelinating Gardner syndrome; familial myelodysplastic syndrome;
mental health condition; immunological disorder; allergy; infertility;
bronchial asthma; Aveilino type eosinophilia; lung disease; deafness;
reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
gastric disorders; pancreatic disease; Schistosoma mansoni infection;
Spinocerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;
Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
antidiabetic; antidiabetic; antinfertility; antinflammatory;
gynaecological; antidiabetic; hepatotropic; virucide; ophthalmological;
gene therapy; FCTR3a; neurestin-like protein; FCTR3f.
XX
OS Homo sapiens.
XX
PN WO200166747-A2.
XX
PD 13-SEP-2001.
XX
PF 05-MAR-2001; 2001WO-US07160.

```
XX 03-MAR-2000; 2000US-186592P.
PR 03-MAR-2000; 2000US-186718P.
PR 06-MAR-2000; 2000US-187293P.
PR 06-MAR-2000; 2000US-187294P.
PR 17-MAR-2000; 2000US-190400P.
PR 07-APR-2000; 2000US-196018P.
PR 03-JAN-2001; 2001US-259548P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Vernet CAM, Fernandes E, Shinkets RA, Herrmann JL, Majumder K;
PI Macdougall J, Mishra V, Mezes PS, Rastelli L;
XX
XX WPI; 2001-596837/67.
XX N-PSDB; AAS14089.
XX
XX Novel polypeptides designated as FCTRX polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states -
XX
XX Claim 1; Page 39; 215pp; English.
XX
XX The invention relates to human FCTRX polypeptides, FCTRI-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukaemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC allolimmune thrombocytopaenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avellino type
CC eosinophilia, lung diseases, reproductive disorders, deafness, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric
XX
XX Query Match 95.7%; Score 1931; DB 22; Length 2724;
Best Local Similarity 98.9%; Pred. No. 2.3e-151;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 10 KPSAEAGRPITPTSSPILLSAQLPSSHNPPVSCOMPLDSTNSHQIMDTNPDEEFSN 69
DB 162 KSDENGRIPTPTSSPILLSAQLPSSHNPPVSCOMPLDSTNSHQIMDTNPDEEFSN 221
QY 70 SYLLRACSGPQQASSGGPPNHHSSQSTLRPLPPPHNHTLSHHSSANSLSNRSLTNRSSQ 129
DB 222 SYLLRACSGPQQASSGGPPNHHSSQSTLRPLPPPHNHTLSHHSSANSLSNRSLTNRSSQ 281
QY 130 IHAPAPNDLATTPESVQLQDSWVLSNSVPLETRHFLFKTSSGTPLESSSGPGPLTS 189
DB 282 IHAPAPNDLATTPESVQLQDSWVLSNSVPLETRHFLFKTSSGTPLESSSGPGPLTS 341
QY 190 GTVYTPPRLPNTTSRAFKLKKPKSKYCSWKCAALSAATAALLAILLAYFIVPWSLK 249
DB 342 GTVYTPPRLPNTTSRAFKLKKPKSKYCSWKCAALSAATAALLAILLAYFIVPWSLK 401
QY 250 NSSIDSGEAEVGRVTVQVPPGFWRSQIHISQFLKFNISLQKDALFGVYIRRGPLPS 309
DB 402 NSSIDSGEAEVGRVTVQVPPGFWRSQIHISQFLKFNISLQKDALFGVYIRRGPLPS 461
QY 310 HAQYDFMERLDGKEKWSVSPRRRSIQTLVONEAVFVQYLDVGLWHLAFYNDGDKEM 369
DB 462 HAQYDFMERLDGKEKWSVSPRRRSIQTLVONEAVFVQYLDVGLWHLAFYNDGDKEM 521
QY 370 VSFNTVVLD 378
| | | | | | | |
```

```
Db 522 VSFNTVVLD 530
RESULT 5
AAU08680
ID AAU08680 standard; Protein; 2733 AA.
XX AC
XX AAU08680;
XX
XX 18-DEC-2001 (first entry)
XX
XX Human FCTR3b polypeptide sequence.
DE
XX
XX Human; FCTRX; myelogenous leukaemia; carcinoma; melanoma; glioma;
KW astrocytoma; congenital neonatal allolimmune thrombocytopaenia; infection;
KW neurological disorder; neurodegenerative disorders; nerve trauma;
KW familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
KW demyelinating Gardner syndrome; familial myelodysplastic syndrome;
KW mental health condition; immunological disorder; allergy; infertility;
KW bronchial asthma; Avellino type eosinophilia; lung disease; deafness;
KW reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
KW desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
KW gastric disorders; pancreatic disease; Schistosoma mansoni infection;
KW Spino cerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
KW Corneal dystrophy-Greenou type I; Corneal dystrophy-lattice type I;
KW Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
KW anti-allergic; antiasthmatic; anti-infectivity; anti-inflammatory;
KW antidiabetic; protozoicide; hepatotropic; virucide; ophthalmological;
KW gynaecological; antifertility; immunostimulant; auditory; haemostatic;
KW gene therapy; FCTR3b; neurestin-like protein.
XX
XX Homo sapiens.
OS
XX WO200166747-A2.
XX
XX 13-SEP-2001.
XX
XX 05-MAR-2001; 2001WO-US07160.
XX
XX 03-MAR-2000; 2000US-186592P.
PR 03-MAR-2000; 2000US-186718P.
PR 06-MAR-2000; 2000US-187293P.
PR 06-MAR-2000; 2000US-187294P.
PR 17-MAR-2000; 2000US-190400P.
PR 07-APR-2000; 2000US-196018P.
PR 03-JAN-2001; 2001US-259548P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Vernet CAM, Fernandes E, Shinkets RA, Herrmann JL, Majumder K;
PI Macdougall J, Mishra V, Mezes PS, Rastelli L;
XX
XX WPI; 2001-596837/67.
XX N-PSDB; AAS14085.
XX
XX Novel polypeptides designated as FCTRX polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states -
XX
XX Claim 1; Page 35-36; 215pp; English.
XX
XX The invention relates to human FCTRX polypeptides, FCTRI-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukaemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC allolimmune thrombocytopaenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avellino type
CC eosinophilia, lung diseases, reproductive disorders, deafness, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric
```

CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
CC infection Spinocerebellar ataxia, Plasmodium falciparum parasitaemia,
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
CC Rels-Bucklers corneal dystrophy. This sequence represents FCtr3b, a
CC neurestin-like protein.
XX
SQ Sequence 2733 AA;

Query Match 95.78; Score 1931; DB 22; Length 2733;
Best Local Similarity 98.98; Pred. No. 2.3e-151;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 10 KPSAEAGRPPTSPSPSLPSAQLPSSHNPPVSCOMPLDSTSHQIMDTNDEEFSN 69
DB 162 KSDDEGRPIPTSPSPSLPSAQLPSSHNPPVSCOMPLDSTSHQIMDTNDEEFSN 221
QY 70 SYLLRACSGQQASSSSGPPNHHQSOTLRPPPHNHTLSHHSSANSLSNLSLTNRSSQ 129
DB 222 SYLLRACSGQQASSSSGPPNHHQSOTLRPPPHNHTLSHHSSANSLSNLSLTNRSSQ 281
QY 130 IHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFTSGSTPLFSSSPGYPLTS 189
DB 282 IHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFTSGSTPLFSSSPGYPLTS 341
QY 190 GTVYTPPRLPLRNTFSKAFKLLKPKSKYCSWKAALSAIAAALLAILLAYFTVPWSLK 249
DB 342 GTVYTPPRLPLRNTFSKAFKLLKPKSKYCSWKAALSAIAAALLAILLAYFTVPWSLK 401
QY 250 NSSIDSGEAEVGRVTOEVPVPPGFWRSQIHISQPFKFNISLGKDALFGVYIRRGPLPS 309
DB 402 NSSIDSGEAEVGRVTOEVPVPPGFWRSQIHISQPFKFNISLGKDALFGVYIRRGPLPS 461
QY 310 HAQYDFMERLDGKEKWSVESPRRRRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEM 369
DB 462 HAQYDFMERLDGKEKWSVESPRRRRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEM 521
QY 370 VSFNTVVLD 378
DB 522 VSFNTVVLD 530

RESULT 6
ABG03234
ID ABG03234 standard; Protein; 865 AA.

XX AC ABG03234;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #3225.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS67421.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity

PS Claim 20; SEQ ID No 33593; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 865 AA;

Query Match 49.28; Score 992.5; DB 22; Length 865;
Best Local Similarity 61.88; Pred. No. 8.1e-74;
Matches 214; Conservative 12; Mismatches 47; Indels 73; Gaps 6;

QY 84 SSGPPNHSOSTLAPPLPPPHNHTLSHHSSANSLSNLSLTNRSSQIHAPAPNDLATTT 143
DB 158 SSTAPSPPTITTRCPITTP-----SANSLSNLSLTNRSSQIHAPAPNDLATTT 207
QY 144 PESVOLQDSWVLSNVPLETRHFLFTSGSTPLFSSSPGYPLTSGTVYTPPRLPLRN 203
DB 208 PESVOLQDSWVLSNVPLETRHFLFTSGSTPLFSSSPGYPLTSGTVYTPPRLPLRN 267
QY 204 TFSKAEKLLKPKSKYCSWKAALSAIAAALLAILLAYFTVPWSLKNSIDSGAE----- 259
DB 268 TFSKAEKLLKPKSKYCSWKAALSAIAAALLAILLAYFTAGPQMERKASSQVKDRRLR 327
QY 260 -----VGRVTVQEPVPPGFWRSQIHISQPFKFNISLGKDALFGVYIRRGILP----- 307
DB 328 EILLFVGNSTVSPFPD--WRQTV-----QEMMGKSKSCFMERHMTLMTWTAGI 375
QY 308 -----PSH-----AOYDFMERLDGKEKWSVESPR 332
DB 376 ISCILOKQKQAWAGERTCPRHKASKQODLRITKKIIVDATYDFMERLDGKEKWSVESPR 435
QY 333 ERRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEMVSFNTVVLD 378
DB 436 ERRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEMVSFNTVVLD 481

RESULT 7

ABP53587

ID ABP53587 standard; Protein; 2721 AA.

XX AC ABP53587;

XX DT 17-DEC-2002 (first entry)

XX DE Human NOVI5b protein SEQ ID NO:38.

XX KW Human; NOVX; cytostatic; neuroprotective; anticonvulsant; cardiovascular;
XX cerebroprotective; nootropic; antidiabetic; antiinflammatory; fungicide;
XX antirheumatic; antiarthritic; immunosuppressive; antiallergic; virucide;
XX antianaemic; antibacterial; protozoacide; antihelminthic; gene therapy;

CC The present invention describes novel human proteins designated NOVX,
 CC where x is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,
 CC cytostatic, anticonvulsant, cerebroprotective, nootropic, cardiovascular,
 CC antidiabetic, antiinflammatory, antirheumatic, antithrombotic, virucide,
 CC immunosuppressive, antiallergic, antianaemic, antibacterial, fungicide,
 CC protozoacide and antihelminthic activities, and can be used in gene
 CC therapy. The NOVX proteins, nucleotides or antibodies can be used in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,
 CC ischaemic cerebrovascular disease, Alzheimer's disease or Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or goitre), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The NOVX proteins can be used as immunogens to
 CC produce antibodies and as vaccines. The NOVX nucleotide sequences may
 CC be used in chromosome mapping, identifying individuals from minute
 CC biological samples (tissue typing), and in forensic identification of a
 CC biological sample. The present sequence represents human NOV15c, which is
 CC located on chromosome 4.
 CC
 CC Sequence 2628 AA;

Query Match 48.2%; Score 972.5; DB 23; Length 2628;

Best Local Similarity 55.1%; Pred. No. 1.7e-71;

Matches 190; Conservative 46; Mismatches 58; Indels 51; Gaps 5;

QY 87 PNNHSSQSLRPLPPPHNHTLSHHSSANSLSNRSLNRSQIHA-PAPAPNDLATPTE 145
 DB 80 PASNOGQSTLQ-PLPPSHRQHSQAHHPSITSLNRSLNRRNQSPAPPALPAELQTTPE 138
 QY 146 SVQLQDSWLSNVPLTETRHFLFKTSGSTPLFSSSGVPLTSGTVTPPPRLPRNTF 205
 DB 139 SVQLQDSWLSNVPLTETRHFLFKTSGSTPLFSSSGVPLTSGTVTPPPRLPRNTL 198
 QY 206 SRKAFKLPKPSKCSKCAALSAAIAALLAILLAYFI-----VPSWL----- 248
 DB 199 SRSAFKKSKSKCSKCAALCAVGVSVLLAILLSYFIAMHLFGLNWLQQTENDFENG 258
 QY 249 -----KNSIDSGEAGVGRVTVQVPPGVFWRWSQIHS 281
 DB 259 KVNSDTMTNTVSLPSGDKGLGGFTQENNTDSGELDIGRAIOIPPGFIWRWSQLFID 318
 QY 282 QPQFLKFNLSLGLKALFGVYIRRGPPSHQAQDFMERLDG-----KEKSVVESPRERS 336
 DB 319 QPQFLKFNLSLQKDALIGYGRKGLPPSHQTQDFVELLDGSLRIAREQSLLETERAGRQ 378
 QY 337 IOTLVQNEAVFOYLDVGLWHLAFYNDGDKEMVSNFTVYLDGTI 381
 DB 379 ARSVSLHEAGFIQYLDGSIWHLAFYNDGKNAEQVSENTIVIESVW 423

RESULT 10

ABP53586

ID 'ABP53586 standard; Protein; 2725 AA.

XX AC ABP53586;

XX AC ABP53586;

DT 17-DEC-2002 (first entry)

XX Human NOV15a protein SEQ ID NO:36.

XX Human; NOVX: cytostatic; neuroprotective; anticonvulsant; cardiovascular;
 KW cerebroprotective; nootropic; antidiabetic; antiinflammatory; fungicide;
 KW antirheumatic; antiarthritic; immunosuppressive; antiallergic; virucide;
 KW antianaemic; antibacterial; protozoacide; antihelminthic; gene therapy;
 KW cancer; leukaemia; lymphoma; melanoma; neurological disorder; epilepsy;
 KW stroke; ischaemic cerebrovascular disease; Alzheimer's disease; allergy;
 KW Pick's disease; vesicular transport disease; cystic fibrosis; goitre;
 KW diabetes mellitus; Grave's disease; gastrointestinal disorder; vaccine;

KW ulcerative colitis; gastric disorder; duodenal disorder; infection;
 KW autoimmune disease; allergic reaction; autoimmune haemolytic anaemia;
 XX rheumatoid arthritis; chromosome 4.
 OS Homo sapiens.
 XX WO200262999-A2.
 XX 15-AUG-2002.
 XX 31-DEC-2001; 2001WO-US49976.
 XX 29-DEC-2000; 2000US-258928P.
 PR 02-JAN-2001; 2001US-259415P.
 PR 04-JAN-2001; 2001US-259785P.
 PR 20-FEB-2001; 2001US-269814P.
 PR 09-MAR-2001; 2001US-279863P.
 PR 29-MAR-2001; 2001US-279832P.
 PR 29-MAR-2001; 2001US-279833P.
 PR 13-APR-2001; 2001US-283889P.
 PR 18-APR-2001; 2001US-284447P.
 PR 25-APR-2001; 2001US-286683P.
 PR 29-MAY-2001; 2001US-294080P.
 PR 16-AUG-2001; 2001US-312915P.
 PR 17-AUG-2001; 2001US-31325P.
 PR 17-SEP-2001; 2001US-322699P.
 PR 26-NOV-2001; 2001US-333350P.
 XX (CURA-) CURAGEN CORP.
 PA SPYTEK KA, LI L, Wolenc AR, Vernet CAM, Eisen A, Liu X;
 PI Malyankar U, Shinkets RA, Tchernev VT, Spaderna SK, Gorman L;
 PI Kekuda R, Patturajan M, Gusev V, Gangolli EA, Guo X, Shenoy S;
 PI Rastelli L, Casman SJ, Boldog F, Burgess CE, Edinger S;
 PI Ellerman K, Gunther E, Smithson G, Millet I, MacDougall JR;
 XX WPI; 2002-732706/79.
 DR N-PSDB; ABQ82343.
 XX
 PT NOVX polypeptides and polynucleotides useful for treating
 PT NOVX-associated disorders, such as cancers, neurological disorders,
 PT disorders of vesicular transport, gastrointestinal disorders, and
 PT autoimmune diseases
 XX
 PS Claim 1; Page 113; 444pp; English.
 XX
 CC The present invention describes novel human proteins designated NOVX,
 CC where x is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,
 CC cytostatic, anticonvulsant, cerebroprotective, nootropic, cardiovascular,
 CC antidiabetic, antiinflammatory, antirheumatic, antithrombotic, virucide,
 CC immunosuppressive, antiallergic, antianaemic, antibacterial, fungicide,
 CC protozoacide and antihelminthic activities, and can be used in gene
 CC therapy. The NOVX proteins, nucleotides or antibodies can be used in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,
 CC ischaemic cerebrovascular disease, Alzheimer's disease or Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or goitre), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The NOVX proteins can be used as immunogens to
 CC produce antibodies and as vaccines. The NOVX nucleotide sequences may
 CC be used in chromosome mapping, identifying individuals from minute
 CC biological samples (tissue typing), and in forensic identification of a
 CC biological sample. The present sequence represents human NOV15a, which is
 CC located on chromosome 4.
 CC
 CC Sequence 2725 AA;

Query Match 47.4%; Score 956.5; DB 23; Length 2725;


```

Db      476 GRGLPSTHTQDFVLLDGRLLTQANSLGTPQSRGTVPPSSSHENGFIQYLDGSIW 535
QY      357 HLAIFYNDGDKKEMVSFNTVVLID 378
Db      536 HLAIFYNDGKESEVVSFLTIAIE 557

RESULT 12
ABG70388
ID ABG70388 standard; Protein; 2769 AA.
AC ABG70388;
XX
DT 05-NOV-2002 (first entry)
DE Human TEN-M4-like protein.
XX
KW Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic pathway modulation; metabolic disorder;
KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;
KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;
KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;
KW memory defect; infertility; congenital heart defect; hair growth;
KW pigmentation disorder; endocrine disorder; respiratory disease; health;
KW gastro-intestinal disease; reproductive; neurological disease;
KW bone marrow transplantation; endocrine disease; allergy; inflammation;
KW neurological disorder; urinary system disorder; age-related disorder;
KW neuropsychiatric disorder; EGF-related protein; SCUBEL; TEN-M4;
KW adipocyte complement-related C1q tumour necrosis factor; out at first;
KW beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;
KW type Ia membrane sushi-containing domain; butyrophilin;
KW type Ia membrane-sushi domain containing; SNP;
KW single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 28 /note= "Asp substituted by Gly as a result of a single
FT nucleotide polymorphism (SNP)"
FT Misc-difference 64 /note= "Val substituted by Ala as a result of a single
FT nucleotide polymorphism (SNP)"
FT Misc-difference 76 /note= "Ala substituted by Thr as a result of a single
FT nucleotide polymorphism (SNP)"
FT FT
XX WO200257453-A2.
XX
XX 25-JUL-2002.
XX
XX 19-DEC-2001; 2001WO-US50331.
XX
XX 19-DEC-2000; 2000US-265704P.
XX
XX 20-DEC-2000; 2000US-257314P.
XX
XX 02-MAY-2001; 2001US-288153P.
XX
XX 29-MAY-2001; 2001US-294075P.
XX
XX 24-JUL-2001; 2001US-307506P.
XX
XX 10-AUG-2001; 2001US-311590P.
XX
XX 10-AUG-2001; 2001US-311613P.
XX
XX 29-AUG-2001; 2001US-315617P.
XX
XX 14-SEP-2001; 2001US-322358P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gangolli EA, Patturajan M, Vernet CAM, Malyankar UM, Kekuda R;
XX Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zerhusen BD, Liu X;
XX Spytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;
XX WPI; 2002-590744/63.
XX N-PSDB; ABS52100.
XX

```

```

PT Novel isolated NOVX polypeptide useful for treating cardiomyopathy,
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or
PT cancer
XX
PS Claim 1; Page 53; 318pp; English.
XX
CC The present invention relates to new NOVX polypeptides. The invention is
CC useful for treating or preventing a NOVX-associated disorder such as
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell
CC signal processing and metabolic pathway modulation in a subject,
CC preferably human. The invention is also useful for treating metabolic
CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune
CC disorders, haematopoietic disorders and various cancers. The molecules of
CC the invention are also useful for treating or preventing cirrhosis,
CC pancreatitis, learning and memory defects, infertility, congenital heart
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,
CC respiratory disease, gastro-intestinal diseases, reproductive, health,
CC neurological diseases, bone marrow transplantation, endocrine diseases,
CC allergy and inflammation, nephrological disorders, urinary system
CC disorders, neuropsychiatric disorders and age-related disorders.
CC The present amino acid sequence represents a NOVX protein of the
CC invention.
XX
SQ Sequence 2769 AA;
Query Match 39.3%; Score 792.5; DB 23; Length 2769;
Best Local Similarity 41.4%; Pred. No. 1.8e-56;
Matches 173; Conservative 49; Mismatches 83; Indels 113; Gaps 10;
QY 69 NSYLLRACSGPQOASSGPPNHHQSQTLRPPPLPP-PHNHTLSHH-SNANSLNRNLSLNR 126
Db 149 NSNLTLDTEHENTETDHPGGLQNHARLRTPPPLSHHTPNQHHAAASINSLNRGFTPR 208
QY 127 RSQIHAPAPND--LATTP-----ESVOLQDSWVLSNVPLETR----- 164
Db 209 SN-----PSPAPTDHSLSGEPAGGAOEPAAHOENLLNSNIPLETRNLGKQPPLGTLQDN 264
QY 165 -----HFLFKTSSGSTPLFSSSGVPLTSGTVYTPPPPLPRNTF 205
Db 265 LIEMDILGASRDGAYSDGHFLFK-PGHSPLFCTTSPGTLTSTVYPPPPPLPRSTF 323
QY 206 SRKAFKLKPKSCYKWCKAALSAIAAALLAILLAYFI-----VPWSL----- 248
Db 324 ARPAPFLNPKPKSCYKWCKAALSAIVISATLVLLAVFVAMHFLGLNWHLPQMEGQWEIT 383
QY 249 -----KNSIDSIGEAEGVRRVT 265
Db 384 EDTASSWPVPTDVSLYPSGGTGLETPDRKGKGTCKPSFFPDSFDSIGETDVGRRAS 443
QY 266 QEVPPGVFWRISOIHISQPOFLKFNISLGDALFGVYVIRGLPPSHAOYDFMERLDCK--- 322
Db 444 QKIPGCTFWRISOVFDHPVHLAFNLSLGAALVGIIYGRKGLPPSHQTFQDFVELLDGRLL 503
QY 323 --EKWSVSVESPRERRSIQTLVQNEAVFYQYLDVGLMHLAFYNDGDKKEMVSFNTVVLID 378
Db 504 TQEARSLGTPQSRGTVPPSSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTIAIE 561

RESULT 13
AAW44817
ID AAW44817 standard; Protein; 768 AA.
XX
XX AC AAW44817;
XX
XX DT 17-AUG-1998 (first entry)
XX
XX DE Human gamma-heretulin.
XX
XX Gamma-heretulin; gamma-HRG; human; autocrine growth factor;
KW breast cancer; MDA-MB-175 cell; diagnosis; gene therapy;
KW ErbB receptor; cell proliferation; cell differentiation;

```

cell survival; neurological disorder; muscular disorder.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Domain 1..560

XX Region /note= "N-terminal domain (Claim 5)"

XX /note= "hydrophobic region, may function as internal signal sequence"

XX Domain 704..748

XX Modified-site /note= "EGF-like domain"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX W09802541-A1.

XX 22-JAN-1998.

XX 08-JUL-1997; 97WO-US11841.

XX 12-JUL-1996; 96US-0021640.

XX (GETH) GENENTECH INC.

XX Schaefer GM, Sliwkowski M;

XX WPI; 1998-110589/10.

XX N-PSDB; AAV19251.

XX DNA encoding gamma-heretulin - used to activate ErbB receptor and to enhance proliferation, differentiation or survival of a cell

XX Claim 3; Fig 1A-C; 81pp; English.

XX This polypeptide comprises human gamma-heretulin (gamma-HRG), a novel member of the heregulin superfamily, that has a unique N-terminal domain not present in previously identified heregulins. Gamma-HRG is secreted by human breast cancer MDA-MB-175 cells. It leads to the formation of a constitutive active receptor complex and stimulates the growth of these cells in an autocrine manner. A claimed method for activating an ErbB receptor comprises contact a cell which expresses the receptor with gamma-HRG. A claimed method for enhancing the proliferation, differentiation or survival of a cell, especially a glial cell or muscle, comprises contacting the cell with gamma-HRG. Antibodies raised against the unique N-terminal domain of gamma-HRG can be used to detect and purify the protein. Antagonists of gamma-HRG can be used to block gamma-HRG activity and expression. The gamma-HRG may be obtained from claimed host cells that comprise a vector containing an isolated gamma HRG nucleic acid (see AAV19251).

XX Sequence 768 AA;

Query Match 39.2%; Score 790.5; DB 19; Length 768;

Best Local Similarity 41.5%; Pred. No. 4.7e-57;

Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

XX 69 NSYLLRACSGPQQAASSGPPHHQSOTLRPLPP-PHNHTLSHHH-SSANSINRSLNTR 126

XX 149 NSNLTLDTEHENTETDRPGGLQNHARLRTPPPLSHAHTPNQHHAAASINRGNFTPR 208

XX 127 RSQIHAPAPND--LATTP-----ESVQLQDSWLNNSVPLETR----- 164

DB 209 SN-----PSPAPTDHSLSGEPPAGGAQEPAAHAQENLLNSNIPLETRNLGKQFFLGLQDN 264

QY 165 -----HLEKTSSTGSGTFLSSSSPGYPLTSGTVTPPPRLPRNTF 205

DB 265 LIEMDILGASRDGAYSDGHELFK-PGGSPLFCTTSPGYPLTSTVSPPPRPLPRSTF 323

QY 206 SRKAFKLKPKSKYCSWKCAALSAIAAALLAILLAYFI-----VPWSL----- 248

DB 324 ARPANLKKPKSKYCNWKAALSAIVISATLVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383

QY 249 -----KNSSIDSEAEVGRRTV 265

DB 384 EDTASSWNPVTDVSLYPSGGTGLETPDRKKGTTGKPSFFPEDSFDSDGEIDVGRRAS 443

QY 266 QEVPPGVFWRQIHSIQPOFLKFNISLKGKDALFGVYIRRGPPSHAOYDFMERLDGK--- 322

DB 444 QKIPPGFWRQSVQFIDHPVHLKFNVSLGKAALVGIYGRKGLPPSHITQDFVELLDGRLL 503

QY 323 --EKWSVVESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDEKEMVSFNTVYL 377

DB 504 TQEARSLGTPRQSRGTVPSPSHETGFIQYLDSDGIWHLAFYNDGKSEWSFLLTAAI 560

RESULT 14

AAV06639

ID AAY06639 standard; Protein; 768 AA.

XX AC AAY06639;

XX DT 26-OCT-1999 (first entry)

XX DE Gamma-heretulin.

XX KW Gamma-heretulin; HRG-alpha; human; ligand; HER2; HER3; HER4;

XX KW receptor; lung surfactant; respiratory distress syndrome;

XX KW emphysema; epithelial growth factor; therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Region 342..363

XX Modified-site /note= "hydrophobic region"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Modified-site /note= "N-glycosylated"

XX Domain /note= "N-glycosylated"

XX Domain 704..748

XX /note= "epidermal growth factor-like domain"

XX W09939729-A2.

XX 12-AUG-1999.

XX 03-FEB-1999; 99WO-US02390.

XX 04-FEB-1998; 98US-0020598.

XX (GETH) GENENTECH INC.

XX (IOWA) UNIV IOWA RES FOUND.

XX Kern JA, Sliwkowski M;

XX WPI; 1999-494213/41.

XX N-PSDB; AAX87705.

XX		XX	Heregulin ligands can be used to induce epithelial cell growth, and
PT		PT	to promote repair and healing of tissue damage or injury
XX		XX	[disclosure; Page 108-111; 120pp; English.
CC	This sequence represents gamma-heregulin (gamma-HRG) deduced from		
CC	gamma-HRG cDNA (see AXA87705). The invention provides HRG ligands,		
CC	Including gamma-HRG, that have affinity for and stimulate HER2,		
CC	HER3 and/or HER4 receptors in autophosphorylation. A new method of		
CC	treating respiratory distress syndrome in humans uses HER2, HER3		
CC	and/or HER4 receptor ligands as epithelial growth factors. A novel		
CC	method of inducing epithelial cell growth and/or proliferation		
CC	comprises contacting a normal epithelial cell which expresses HER2,		
CC	HER3 and/or HER4 receptors with an isolated ligand which activates		
CC	HER2, HER3, HER4 receptors or their combination. Also claimed are		
CC	methods of increasing lung surfactant protein A, or of treating		
CC	chronic obstructive pulmonary disease, respiratory distress or		
CC	pneumysena, by administering an effective amount of an isolated HER		
CC	ligand to a patient.		
XX		XX	
SQ	Sequence 768 AA;		
	Query Match 39.2%; Score 790.5; DB 20; Length 768;		
	Best Local Similarity 41.5%; Pred. No. 4.7e-57;		
	Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;		
QY	69 NSYLLRACSGPOCASSGGPPNHHQSQTLPRLPP-PHNHTLSHHH--SSANSLNRNSLTNR 126		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	149 NSNLTTDTEHEWETEDHPGGLGNHAKLRTPPPLSHAHTPNQHHAASINSLNRNGFTPR 208		
QY	127 RSQIHAPAPND--LATTP-----ESVOLQDSVLNNSVPLETR----- 164		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	209 SN----PSPAPTDHLSLGSGEPPAGGAQEPAHAQNWLNSNIPLETRNLKQPFLGTLDQN 264		
QY	165 -----HFLEKTSSTGPLFSSSSPGYPLTSGVTYTPPRLLPRTF 205		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	265 LIEMDILGASRHHDGAYSDDGHLEFK-PGGTSPLECTTSPGYPLTSGVTYTPPRLPRTF 323		
QY	206 SRKAFLKPKSKYCWNKCAALSATAALLAILAYFI-----VPWSL----- 248		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	324 ARPFWLKPKSKYNWKCAALSIVAIVTLILLAYFAMHLFGLNWHLOPMEGQMYEIT 383		
QY	249 -----KNSSIDSGEAEEVGRRVT 265		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	384 EDTASSMPVTDVSLYPSGTGLETPDRKGKTECKPSSFPPEDSFIDSGEIDVGRAS 443		
QY	266 QEYPGVFWRSQTHISQOFLKENISLGKDALFGVIYIRGLPPSHQAQYDFMERLDGK--- 322		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	444 QKIPTGFWRSQVFIDHPVHLKFENSVLGLKAALVGIIYRKGLPPSHTQDFVELLDGRLL 503		
QY	323 --EKSWVESPREERSIQTLVQNEAVFYQLDVGLHHLAFYNDGDKEMYSFNTVVL 377		
DB	: : : : : : : : : : : : : : : : : : : : : : : : :		
	504 TQEARSLGEGTPROSGRTVPVPSSETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAT 560		
RESULT 15			
AAY71176			
ID	AAY71176 standard; Protein; 768 AA.		
XX			
AC	AAY71176;		
XX			
DT	21-SEP-2000 (first entry)		
XX			
DE	Human Heregulin variant, gamma-HRG protein.		
XX			
KW	Heregulin; variant; gamma-HRG; human; inner-ear-supporting cell;		
KW	activator; HER2 receptor; HER3 receptor; HER-2 oncogene; chromosome 8p;		
KW	hair cell related hearing disorder; ototoxic injury; tissue damage;		
KW	acoustic assault; degenerative hearing loss; balance impairment;		
KW	treatment; surgical injury; physical injury; inner ear disorder.		
OS	Hom sapiens.		

```

      ||||| |::||| ::||| ||||| ||||| ||||| ||||| |||||
Db 265 LIEMDILGASRHDGAYSDCHELFK--PGCTSPLECTTSPGYPLTSTSTVYSPPPRPLPRSTF 323
QY 206 SRKAFKLRKPSKYCSWKCAALSATAAALLLAILLAYFI-----VPWSL----- 248
      :|| ||||| ||||| ||||| : ||||| : |
Db 324 ARPAFNLRKPSKYCNWKAALSIAVISATVILLAYFVAMHLFGLNHLQPMEGOMYEIT 383
QY 249 -----KNSIDSGEAEVGRVVT 265
      :|| ||||| ||||| :
Db 384 EDTASSNPVPTDVSLYPSGGTGLETPDRKGKTTEGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSQIHISOPQELKENISLGDALFGVYIRRGLPSPSHAOYDEMERLDCK--- 322
      |::||| ||||| : | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 QRIPPGTFRSQVFIIDHPVHLKENVSLGKAALVGIYGRKGLPPSHTQDFVELLDGRLL 503
QY 323 --EKWSVYESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGDKEMVSFNTVVL 377
      | : :||| : | : ||||| ||||| ||||| : ||||| :
Db 504 TQEARSLGTPRQSRGTVPSPSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTITAI 560
```

Search completed: August 26, 2003, 14:31:00
Job time : 60 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:28 ; Search time 19 Seconds
(without alignments)
1928.435 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHAANKGRPSAAGRP.....NDGKDKEMVSPNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	853	42.3	2825	2 T14271	Doc4 protein, stre
2	205.5	10.2	2406	2 A54148	od2 protein - fruit
3	186.5	9.2	2515	2 S47008	tenascin-like prot
4	156.5	7.8	528	2 I47141	gastric mucin (clo
5	150	7.4	782	2 A61625	tenascin-like prot
6	141	7.0	2531	2 T16743	hypothetical prote
7	134.5	6.7	633	2 S62057	proline-rich prote
8	132.5	6.6	2715	2 T13049	eyelid - fruit fly
9	132	6.5	1268	2 T31420	C-terminal domain-
10	131	6.5	1892	2 T18314	hypothetical prote
11	129.5	6.4	731	2 T04455	hypothetical prote
12	129	6.4	383	2 T46707	proteophosphoglyc
13	129	6.4	1456	2 T01397	LTR gag/pol polypr
14	128	6.3	356	2 A96826	T8K14.10 (imported
15	127.5	6.3	554	2 T49833	related to Vea pro
16	127.5	6.3	1420	2 T37781	probable cytoskele
17	126.5	6.3	720	2 J01676	AB13 protein - Ara
18	126	6.2	620	2 D67733	hydroxyproline-ric
19	126	6.2	710	2 D36728	hypothetical prote
20	126	6.2	2232	2 T34434	hypothetical prote
21	124.5	6.2	731	2 B86369	hypothetical prote
22	123.5	6.1	317	2 S55316	mucin (clone PGM-2
23	123.5	6.1	1097	2 T13033	cyclin T - fruit f
24	123.5	6.1	1140	2 D8690	protein F4H10.3 (
25	123	6.1	193	2 C86334	hypothetical prote
26	123	6.1	490	2 S52830	HMS1 protein - yea
27	122.5	6.1	968	2 S46992	protein p130 - rat
28	122.5	6.1	1680	2 T41628	probable transcrip
29	121.5	6.0	760	2 F86387	probable Pto kinas

RESULT 1

T14271

Doc4 protein, stress-induced - mouse

N:Alternate names: od2 protein homolog

C:Species: Mus musculus (house mouse)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14271

R:Wang, X.Z.; Kuroda, M.; Sok, J.; Batchvarova, N.; Kimmel, R.; Chung, P.; Zinszner,

EMBO J. 17, 3619-3630, 1998

A:Title: Identification of novel stress-induced genes downstream of chop.

A:Reference number: Z17951; MUID:98315054; PMID:9649432

A:Accession: T14271

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2825 <W>

A:Cross-references: EMBL:AF059485; NID:g3170614; PID:g3170615; PIDN:AAC31807.1

C:Genetics:

A:Gene: Doc4

Query Match	42.3%	Score	853;	DB 2;	Length	2825;			
Best Local Similarity	40.1%	Pred. No.	2.6e-51;						
Matches	193;	Conservative	55;	Mismatches	111;	Indels	122;	Gaps	12;
QY	12	SAEAGRIPTTSPSLLPSAQLPSSHNPPVS-CQPLLDNSTSHQIMDTPDEEFSNPS	70						
Db	161	NTEGAPL-HCSSASSTPIEQSPSPPANESQRRLLGNQVAQPTPDSDEEFVNS	219						
QY	71	YLLRACSGPQOQASSGPP-----NHQSQTLRPLPPPHNHTLSHH--SSANSLRNSLTN	125						
Db	220	FLVKSGSASLGVAANDHPSSLQNHPRLT--PPPLPHAHTPNQHAASINSLNRGNFT	277						
QY	126	RRSQIHAPAPND-----LATTPESVOLQDSWVLSNVPLETR-----	164						
Db	278	RSN-----PSPAPTDHLSGEPAGSAQEPHTAQNWVLSNKIPVETRLGKQPF	333						
QY	165	-----HFLKTSSTGTPFLSSSPGYPLTSGTYVTPPPRLPLPRNT	204						
Db	334	NLIEMDIFSASRRDGAYSQGHFFFK-PGQTSPLCTTSPGYPLTSTVYSPRPRPLRST	392						
QY	205	FSRAFKLKKPSKYCSWKAALSAIAAALLAILLAYFI-----VPSNL-----	248						
Db	393	FSRPAFLKPKSKYCNWKNALSAILSATVILLAYFVAMHLFGLNHLNLPMEGQOMQY	452						
QY	249	-----KSSSIDSGAEVGR	262						
Db	453	EITEDTASWPVPTDVSLYPSGGTGLETPDRKKGAAEGKPSLFPEDSPIDSGEIDVGR	512						
QY	263	RVTQVPPGVFWBSQIHISQPFKFNISLKGKDALFGYVIRGLPPSHQAQDFMERLDGK	322						
Db	513	RASQIPPGTFRSQVFDHPVHLKFNVLGKALVGIYGRKGLPSPHTQLDFVELLDGR	572						
QY	323	-----EKWSVSESPRRERSITQLYQNEAVFYQYLDVGLWHLAFYNDGDKREMSFNTVVL	377						

Query Match 6.7%; Score 134.5; DB 2; Length 633;
Best Local Similarity 25.1%; Pred. No. 0.074;
Matches 55; Conservative 24; Mismatches 83; Indels 57; Gaps 8;
QY 18 PIPPTSSPSLLPSAQLPSSHN-----PPVSCQ-----MPLLDSTSHQ 56
DB 264 PSQSQNPNPPFPIPEIPSTQSATNPPFPVQOQFNOAPSMGIPQONRPLPQPNRNNRP 323
QY 57 IMDTNPDEEFPNSYLLRACSGPQQASSGPPNNHHSQSTLRPLPPPHNHTLSHHSSAN 116
DB 324 VPPPPPMRTTEGSGVRLPAPPPPPRRGAPP-----PPPHRVTS---NTLN 368
QY 117 SLNRSLNRRSIOIHAPAPNDLATPE---SVQLQDSWVLNSVPLETRHFLFTSSG 173
DB 369 SAGNSLLQATGRRGPAPPPPPRASRPTNVTMOQNPOQYNNRNP-----FGYTNSN 423
QY 174 STPLFSSSPGYPLTSGTVTP-----PPRLPNT 204
DB 424 -----MSSPPPPPTVFNLTQMTAATGQPAVPLPQNT 457
RESULT 8
T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TRE>
A:Cross-references: EMBL:AF053091; NID:G2981220; PID:G2981221; PIDN:AAC06254.1
A:Gene: eld
A:Function:
C:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding
Query Match 6.6%; Score 132.5; DB 2; Length 2715;
Best Local Similarity 24.3%; Pred. No. 0.06;
Matches 64; Conservative 21; Mismatches 97; Indels 79; Gaps 13;
QY 8 GRKPSAAGRPPTSSPSLLPSAQLP-----SSHNPPEVS-----COMPLDST 53
DB 413 GSSPSPGSGHPLPPASPHVHPVLPQQPPPPHVSAGGPPSPSGHAPSPSPQASPS 472
QY 54 SHQ-IMDTPNDEEFS-----PN-SYLLRACSGP----- 79
DB 473 PHQELIGQNSDSSSGAHSGMGSGPPTPNQOVNRPTPTSGSGSRMSPAVAQNH 532
QY 80 -----QQASSGPPNNHHSQSTLRPLPPH-----NHTLSHHSSANSNLRSLTN 125
DB 533 ISRPASNQSSSGPMQPPVAGGPPMPHPGMPGPPQQQSQSQASNS--ASSASN 590
QY 126 RRSQIHAPAP-----NDLATTPEVSQLODSWVLNSVPLETR-HFLFTSSGSTPLFS 180
DB 591 SPOQTTPPPAPPPNQGNHNNATPPPPQ-----GAAGGYPMPPHMHGKMGK---PQSP 643
QY 181 SSPGYPLTSGTVTP---PPR 198
DB 644 GAQGYPPQPPQYPPGNYPPR 664
RESULT 9
T13420
C-terminal domain-binding protein ra8 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C:Accession: T13420

R:Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; C
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts
A:Reference number: Z21024; MUID:96293459; PMID:8692929
A:Accession: T31420
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1268 <YUR>
A:Cross-references: EMBL:U49055; NID:G1438529; PID:G1438530; PIDN:AAC52656.1
A:Experimental source: hippocampus
Query Match 6.5%; Score 132; DB 2; Length 1268;
Best Local Similarity 29.2%; Pred. No. 0.27;
Matches 63; Conservative 22; Mismatches 85; Indels 46; Gaps 14;
QY 11 PSAACRPIPTSSPSLLPS-AOLPSSHNPVSC-OMPLDSTNTSHQIMDTPNDEEFS 68
DB 778 PSCESTRPVIPSDIPSAAPMAQPGASNTSGILQVORPNVSSNS--EILGVSPAN--VS 833
QY 69 NSYLLRACSGPQQASSG-----PPNHHSQSTLRPLPP--PNHHTLSHHSSANSNLRN 121
DB 834 NSAAINGAOPPPNNLNSGILGIQPPNVSSGSLGLVLPPLPN-----SGLVGLQP 886
QY 122 SLNRRSQIHAPAP-APNDLATTPESVQLQDSWVLNSVP-----LETRHFLFTSSGSTP 176
DB 887 NVTNPAGLLGTQPIGPQNL--PPLTIPAQ-----RMPALPMLDIRPGLIAQAFC--P 935
QY 177 LFSSSPGYPLTSGTVTPPPRL-----PRNTF 205
DB 936 RFPLLQPGIPQPG-----IPPPSVLDAALHPPPRGPF 968
RESULT 10
T18314
hypothetical protein L7610.4 - Leishmania major
C:Species: Leishmania major
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18314
R:Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.; Ivans, A
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18876
A:Accession: T18314
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1892
A:Cross-references: EMBL:AL034356; NID:el371878; PID:el371559; PIDN:CAA22240.1
C:Genetics:
A:Note: L7610.4
Query Match 6.5%; Score 131; DB 2; Length 1892;
Best Local Similarity 20.3%; Pred. No. 0.53;
Matches 88; Conservative 57; Mismatches 169; Indels 120; Gaps 19;
QY 12 SAAGAPPIPTSSP-----SLLPSAQLPSSHNP-PPVSCOMPLDSTNTSHQIMDTPNDEEFS 67
DB 625 SATVAAPITKTDPYAGHTAPDGEPRHLVMPMPPIQRPYAAT-----EEGA 673
QY 68 PNYLLRACSGPQQASSGPPNNHHSQSTLRPLPPPHNHTLSHHSSANSNLRSLNRR 127
DB 674 PR-FSVRKVTAPQESDAPSPH-----SPPAEHTLLSRGAADA-GEAAAKETR 721
QY 128 SQIHAPAPNDLATTT-----PESVQLQ-----DSWVLNSVPLETR 164
DB 722 EQLSAAKEAVTAMTTAGVQSSKKPQOLQHEPDGSPNGDDVDELLEADLLIMQSRPQSSR 781
QY 165 HFLFTSSGSTPLFSSSPGYP-----LTSGTVYT-----PPRLLP 201
DB 782 H---GAAGSSPIQSIHSPQSPAVECFSDAPRSVHATVGTSETTEQQQQQORQHPOLRP 838
QY 202 RNTFSRKAFKLRKPSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKSSIDSGAEV 261
DB 839 -TSLGRGAVASDSRCAAEKAEELTPPPPAQL-----PTAMD-----DITERDVG 884

QY 262 RRTVOEPPGVFMRQIHSOQFLKFNLSLGDALFGVYIRRL-PPSHAYQDFM---- 316
Db 895 EQ-----SNVYIRGHMMNTLPQKMDAPRCGGSTAVHGA--TEGLEPPQDVAVDVYMGTA 936
QY 317 ERLDGKE-----KWSVVEGPRRRSTQTLVQNEAVFQVLDVGLWHLA 359
Db 937 ELREGKEALCGPSTGADVGPKTLQEAADTSEPTGASAPLVLRERYVYVDPLEASMDAA 996
QY 360 FYNQKQKEMVSN 373
Db 997 YMLDTLEQQYVRIN 1010

RESULT 11

T04455

Hypothetical protein F4D11.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 04-Mar-2000

C:Accession: T04455

R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Hohelsel, J.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15360

A:Accession: T04455

A:Molecule type: DNA

A:Residues: 1-731 <BEV>

A:Cross-references: EMBL:AL022537

A:Experimental source: cultivar Columbia; BAC clone F4D11

C:Genetics:

A:Map position: 4

A:Introns: 326/1; 476/1; 505/1; 528/3; 557/2; 606/3; 660/3

A:Note: F4D11.90

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 6.4%; Score 129.5; DB 2; Length 731;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 93; Conservative 33; Mismatches 148; Indels 121; Gaps 17;

QY 11 PSAAEGRPIPTSSPSLLPSLPSQALPSS-HNPPVSCOMPLDSDNTSHQIMDT-----NPDEE 65
Db 21 PPADS---VPDTSSPPAPPLSLPPLPSPPPLPSAPPLSAPTASPPPLPVESPPSPPIE 77
QY 66 FSPNSYLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPNHTLSHHSSANSL 118
Db 78 -SPPPLLESPPPLESPSPSPVSPVSPSPPLPFLPAKPPSPSPSPSPSPSPSPSPSPSPSP 136
QY 119 N-----RNSLTNRSSQIHAPAPA-----PNDLATTTPESVOL 149
Db 137 SPPPSRLPSESTPPVNTASPPPPSPRRS---GPKPSPPPIINSPPNPSPTSLPET 193
QY 150 QDSWVLSNVPLETRHFLFKTSSGSTPLFSSS-----SPGYPLSGTYVTP----- 195
Db 194 SPP-----PKPPLSTTPF-----PSSSTPPPKSPAAVTLPPFGPAGQLPDGTVAPPIGPVI 245
QY 196 -----PPRLPRTFRKAFKLPKSKYCSWKCAAL--SAIAAALLAIL 239
Db 246 EPKTSFASISPGTQPLVPKS-----LPVITSYHRSNAGFLGGVIVGALLLILG 297
QY 240 AYFIVPWSLKNSSIDSGAEVGRRTQVEP---PGVFRSQIH-ISOQFLKFNISLGD 295
Db 298 LLEFVYRATRNRNNNSAAHQSKTPSKVQHRGNGAGTNAHVITMP----- 345
QY 296 ALFGVYIRGLPPSHQAQDFMERLDGKEKWSVES 330
Db 346 -----PPIHAKYISGGCDTKENNSVAKN 369

RESULT 12

T46707

proteophosphoglycan, membrane-associated [imported] - Leishmania major (fragment)

C:Species: Leishmania major

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 11-May-2000

C:Accession: T46707

R:Ilg, T.; Montgomery, J.; Stierhof, Y.D.; Handman, E.

J. Biol. Chem. 274, 31410-31420, 1999
A:Title: Molecular cloning and characterization of a novel repeat-containing Leishman
a1dylinositol anchor.
A:Reference number: Z23133; MUID:20002668; PMID:10531342
A:Accession: T46707
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-383 <ILG>
A:Cross-references: EMBL:AJ243460; PIDN:CAB46680.1
A:Experimental source: strain LRC-1137
C:Genetics:
A:Gene: pp91
C:Function:
A:Description: may be important for direct host-parasite interactions

Query Match 6.4%; Score 129; DB 2; Length 383;
Best Local Similarity 25.0%; Pred. No. 0.095;
Matches 51; Conservative 34; Mismatches 91; Indels 28; Gaps 4;

QY 4 AANKGRKPSAEAGRIPTTSS-PSLLPSAQLPSSHNPPVSCOMPLDSDNTSHQIMDTNP 62
Db 74 SASSSAPSSSSSAPSSASSSAPSSSSSAPSSASSSAPSSSSSAPSSASSSS 124
QY 63 DEEFPNSYLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPNHTLSHHSSANSLNRNS 122
Db 125 ----APSS-----SSAPSSASSSSAPSSSSSAPSSASSSSAPSSSSSAPSSASSSS 176
QY 123 LTNRRSQIHAPAPAPNDLATTTPESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSS 182
Db 177 SSSAPSSASSSSAPSSSSSAPSSAS 250
QY 183 PGYPLTSGTYVTPPPRLPLRNTFS 206
Db 227 PSSSTTTTMDTPDPVLPSSSSS 250

RESULT 13

T01397

LIR gag/pol polyprotein homolog T419.16 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Nov-1999

C:Accession: T01397

R:Parnell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.; Habermann, K.; Schutz, K.;
submitted to the EMBL Data Library, May 1998

A:Description: Genomic sequence of BAC T419 from Arabidopsis thaliana, chromosome IV,

A:Reference number: Z14314

A:Accession: T01397

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1456 <PAR>

A:Cross-references: EMBL:AF069442; NID:g3242970; PID:g3924609

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 129/1

A:Note: T419.16

C:Superfamily: retrovirus-related polyprotein

Query Match 6.4%; Score 129; DB 2; Length 1456;
Best Local Similarity 22.0%; Pred. No. 0.52;
Matches 87; Conservative 42; Mismatches 147; Indels 120; Gaps 17;

QY 11 PSAAEGRPIPTTSSPSLLPSAQLPSSHNPPVSCOMPLDSDNTSHQIMDTPDEEFPNS 70
Db 758 PTTPLVLPAPPCLGPHLDTSPRPPSPSP---LCITQVSSSNL-----PSS 800
QY 71 YLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPNHTLSHHSSANSLNRNSLTNRROI 130
Db 801 ----SISPSSEPTAPSHNGPQTAQP-----HOTQNS-NSNSPILNPNP 842
QY 131 HAPAP-APNDLATTTPESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGVPPLTS 189
Db 843 NSPSPNSPNQNSPLPQS-----PISSPHI-----PTPSTSISEPNSSPS 882

190 GTVYTPPPRLPR-----NTFS---RFAFKLKPSSKYCKWCAALSAIAAALL 234
883 STSTPLPVPVLPAPPIQVNAQAPVNTSHMATRAKDGIRKPNKYSY---ATSLAENSEP 939
235 LAILLAYFVFW-SLNKSSIDSEAEVGRVTOEVP-----GVFWRSQTHISQPF 285
940 RTAIQAMKDRWRQAMGSEIN---AQIGNHTDVLVPPPPSVTIVGCRW----- 985
286 LKFNISLGDALFGVYIRGLPSPHAQYDFMERLDGKEKWSVYESPRERSIOTLVONEA 345
986 -FTFKFNSDGSNLRKARLVAKG---YNQPCGLDVAETFSPIKSTSRIVLGVAVDRS 1041
346 VFQYLDVGLHMLAFYNDGDKEMVSEFNTVVLDTGI 381
1042 WPIRLQLDV-----NNAFLQGLT 1058

RESULT 14
A96826
T8K14.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96826
R:Neologs, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maitl, R.; Marzital,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96826
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <STO>
A:Cross-references: GB:AE005173; NID:g4835761; PIDN:AAD30228.1; GSPDB:GN00141
C:Genetics:
A:Gene: T8K14.10
A:Map position: 1

Query Match 6.3%; Score 128; DB 2; Length 356;
Best Local Similarity 27.7%; Pred. No. 0.1;
Matches 54; Conservative 17; Mismatches 76; Indels 48; Gaps 9;
QY 19 IPTTSPSLPSAQLPSSINPP--PVSCOMPLDSTWTHQIMDTNPDDEFSPNS--YLLR 74
DB 49 LPPLSVPGNAP----PFCINPNTPESSYPGLSPPPGPTITLNPDPSSSNPNPNPPE 104
QY 75 ACSGPOQASGSSGPPNHSQSTLRPLPPPHNHTLSHHSSANSLSLNRSLNRSQIHAPA 134
DB 105 SSSNPNPPSSSNPNPNPPVTVNPP-----ESSNPNPPDSSSNPNPNPNPPE 155
QY 135 PA--PNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSGSTPLFSSSSGYPPLTSGTV 192
DB 156 SSSNPNPPVTVP-----NPP-----ESSNPNPPSSSNPNPNPPII--IP 192
QY 193 YTP-----PPRLLP 201
DB 193 YPPESSSNPNPEIVP 207

RESULT 15
T49833
related to VeA protein [imported] - Neurospora crassa
N:Alternate names: protein B24H17.190
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49833
R:Schutte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-554 <SCH>
A:Cross-references: EMBL:AL356815; GSPDB:GN001116; NCSP:B24H17.190
A:Experimental source: BAC clone B24H17; strain OR74A
C:Genetics:
A:Gene: NCSP:B24H17.190
A:Map position: 6
A:Introns: 56/1

Query Match 6.3%; Score 127.5; DB 2; Length 554;
Best Local Similarity 26.0%; Pred. No. 0.19;
Matches 66; Conservative 19; Mismatches 96; Indels 73; Gaps 14;
QY 5 ANKGRKPSAAGR-PIPTTSSPSLLPSAQLPSSHNPPVSCOMP-----LLDSNTSHQ 56
DB 272 AYQRSMGSGSTERTPYSSISDPQRPS---MADYPPQYAAQTPTSGHGLGFLGGNTHQ 327
QY 57 IMDTNPDEEPS-PNSYLLLRACSGPQQAASSGPPNHHSQSTLRPLPPPHNHTL-SHHSS 114
DB 328 YPAQPPQPSAQPHS---VPPSFVYPTSORAPYQHQPSSYPPP-PPPHQPIFQSEHHTS 382
QY 115 ANSLNRSLNTRRSQIHA-----PAPANDLATTPEVOLQD 151
DB 383 RTYAPINP-ASRHDLSHQSTKQYTLPLSEAVSPTQPHHQHPSIAPHRLPVT----- 433
QY 152 SWVLNSVPLETRHFLFKTSGSTPLFSSS---SPGYP-----LFTSGTVYT---P 195
DB 434 -----ALPPLQVDRF--SSASHNQHMPVSPSNMAAPPYPRAYSVSNSGGLTSAGYNQLPP 487
QY 196 PPRLLPNTFSRKA 209
DB 488 PPPPPFQVAGSKRA 501

Search completed: August 26, 2003, 14:32:27
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:28:28 ; Search time 15 seconds
(without alignments)
1194.478 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHANKGRKPSAAGRP.....NDGRDKEMVSPNTVLDGTI 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134.5	6.7	633	1 LA17_YEAST	Q12446 saccharomyc
2	134	6.6	1822	1 ZAP3_HUMAN	P49750 homo sapien
3	126.5	6.3	720	1 AB13_ARATH	Q01593 arabidopsis
4	126	6.2	620	1 EXTN_TOBAC	P13983 nicotiana t
5	125	6.2	852	1 WS14_HUMAN	Q9np71 homo sapien
6	124.5	6.2	827	1 M4K1_MOUSE	P70218 mus musculu
7	123.5	6.1	1097	1 MCT3_DROME	O96433 drosophila
8	123	6.1	490	1 COT3_YEAST	P54785 saccharomyc
9	123	6.1	1157	1 BCL1_YEAST	P47068 saccharomyc
10	121	6.0	1509	1 GSR1_HUMAN	Q9nzm4 homo sapien
11	121	6.0	5179	1 MUC2_HUMAN	Q02817 homo sapien
12	120.5	6.0	701	1 CG1_HUMAN	Q13495 homo sapien
13	120.5	6.0	968	1 BCAL_RAT	Q63767 rattus norv
14	119.5	5.9	649	1 V70K_EPMV	P20129 eggplant mo
15	117.5	5.8	894	1 M4K3_HUMAN	Q84v78 homo sapien
16	117.5	5.8	1367	1 AMYH_YEAST	P08640 saccharomyc
17	117	5.8	860	1 ARCA_PENRO	Q13508 penicillium
18	117	5.8	2471	1 NTC2_HUMAN	Q04721 homo sapien
19	116.5	5.8	2556	1 NTC1_HUMAN	P46531 homo sapien
20	116	5.8	788	1 PCAP_HUMAN	Q96rn5 homo sapien
21	115.5	5.7	1003	1 MBD6_HUMAN	Q96dn6 homo sapien
22	115	5.7	684	1 RPL1_HUMAN	P78424 homo sapien
23	115	5.7	833	1 M4K1_MOUSE	Q92918 homo sapien
24	115	5.7	2471	1 NTC2_RAT	Q9qwt3 rattus norv
25	114.5	5.7	413	1 ESC1_SCHPO	Q04635 schizosacch
26	114.5	5.7	5085	1 PCLO_RAT	Q91ks6 rattus norv
27	114	5.7	1006	1 P5PA_HUMAN	Q15735 homo sapien
28	114	5.7	3421	1 TEGU_HSVB	P28955 equine herp
29	113.5	5.6	513	1 M113_MOUSE	Q8bgt6 mus musculu
30	113.5	5.6	676	1 MUC1_MESAU	Q60528 mesocricetu
31	113.5	5.6	1132	1 YK5_YEAST	P34250 saccharomyc
32	113.5	5.6	1183	1 DRPL_RAT	P54258 rattus norv
33	113	5.6	890	1 HCNI1_HUMAN	O60741 homo sapien

RESULT 1

ID	LA17_YEAST	STANDARD;	PRT;	633 AA.
AC	Q12446;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Proline-rich protein LAS17.			
GN	LAS17 OR YOR181W.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Toh-E A.;			
RL	Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Hughes B., Pohl T.M.;			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: TO S.POMBE WSP1.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; D78487; BAAL1386.1; -.			
DR	EMBL; 275089; CAA99390.1; -.			
DR	PIR; S62057; S62057.			
DR	SGD; S0005707; LAS17.			
DR	GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.			
DR	GO; GO:0005737; C:cytoplasm; IDA.			
DR	GO; GO:0008092; F:cytoskeletal protein binding activity; IDA.			
DR	GO; GO:0007015; P:actin filament organization; IDA.			
DR	GO; GO:0008154; P:actin polymerization and/or depolymerization; IDA.			
DR	GO; GO:0016288; P:cytokinesis; IMP.			
DR	GO; GO:0006897; P:endocytosis; IMP.			
DR	GO; GO:0007121; P:polar budding; IMP.			
DR	GO; GO:0008970; P:response to osmotic stress; IMP.			
DR	InterPro; IPR000697; EVH1.			
DR	InterPro; IPR001960; WH1.			
DR	InterPro; IPR003124; WH2.			
DR	Pfam; PF00568; WH1; 1.			
DR	Pfam; PF02205; WH2; 1.			
DR	SMART; SM00461; WH1; 1.			
DR	SMART; SM00246; WH2; 1.			
FT	DOMAIN 185 190			POLY-PRO.
FT	DOMAIN 323 329			POLY-PRO.
FT	DOMAIN 342 348			POLY-PRO.
FT	DOMAIN 352 358			POLY-PRO.
FT	DOMAIN 385 391			POLY-PRO.

O35449 mus musculu
P37370 saccharomyc
Q92412 rattus norv
P46938 mus musculu
Q924h2 mus musculu
Q9wub5 rattus norv
Q8wxx7 homo sapien
P10358 turnip yell
P28477 turnip yell
P06599 daucus caro
P06601 drosophila
Q99mz3 mus musculu

ALIGNMENTS

FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 4488355563AA2645 CRC64;
 Query Match 6.7%; Score 134.5; DB 1; Length 633;
 Best Local Similarity 25.1%; Pred. No. 0.051;
 Matches 55; Conservative 24; Mismatches 83; Indels 57; Gaps 8;
 QY 18 PIPPTSPSLPSAQLPSSHN-----PPVSCO-----MPLDLSNTSHQ 56
 DB 264 PSQFQSNPPFPFPEIPSTQSATNPFPPVPOQFNOAFSGMGPQONRPLQPLPNNRP 323
 QY 57 IMDTNPDEEFPSPNLLRACSGPQOASSGPPNNHSSQSTLRPLPPPHNHTLSHHSSAN 116
 DB 324 VPPPPPMRTTTEGSGVRLPAPPPRRGPAPP-----PPPHRHVTS---NTLN 368
 QY 117 SLNRNSLTNRSHIHAPAPNDLATPE---SVQLQDSWLVNSVPLETRHFLFKTSSG 173
 DB 369 SAGGNLLQATGRRGPAPPPPPRPRASRPTPNVTMOONPQOYNNRP-----EGYQNTSN 423
 QY 174 STPLFSSSPGYPILTGTGVTP-----PPRLPNT 204
 DB 424 -----MSSPPPPPVTFNTLPQMTAATGQPAVPLPQNT 457

RESULT 2

ZAP3_HUMAN STANDARD; PRT; 1822 AA.
 AC P49750; P49752; Q9P1V7;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Nuclear protein ZAP3 (ZAP113).
 GN ZAP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bowen L., Madan A., Qin S., Abbasi N., Baradaran L., Birditt B.,
 RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
 RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.,
 RA "Sequencing of human chromosome 14q24.3 region."
 RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=95319502; PubMed=7596406;
 RA Sherrington R., Rogaeve E.I., Liang Y., Rogaeve E.A., Levesque G.,
 RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
 RA Poncin J.-F., Brunl A.C., Montesi M.P., Sorbi S., Rainero I.,
 RA Pinesau L., Nee L., Chumakov I., Pollen D., Brookes A.,
 RA Sansone P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,
 RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
 RA Rommens J.M., St George-Hyslop P.H.;
 RT "Cloning of a gene bearing missense mutations in early-onset familial
 RT Alzheimer's disease."
 RL Nature 375:754-760(1995).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1661.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AC007956; AAF61275.1; -.
 DR EMBL; L40403; AAC42008.1; ALT_FRAME.
 DR EMBL; L40400; AAC42006.1; -.
 KW Nuclear protein.
 FT DOMAIN 15 205 PRO-RICH.
 FT DOMAIN 382 430 GLN-RICH.
 FT DOMAIN 807 1209 ARG-RICH.
 FT DOMAIN 1488 1577 ARG-RICH.
 FT CONFLICT 621 621 P -> S (IN REF. 2).
 FT CONFLICT 1404 1404 T -> I (IN REF. 2).
 FT CONFLICT 1821 1821 K -> E (IN REF. 2).
 SQ SEQUENCE 1822 AA; 204947 MW; 8E6C83FE540C7D2 CRC64;
 Query Match 6.6%; Score 134; DB 1; Length 1822;
 Best Local Similarity 26.6%; Pred. No. 0.2;
 Matches 53; Conservative 22; Mismatches 92; Indels 32; Gaps 7;
 QY 20 PPTSSPSPSLPSAQLPSSHNPPPV--COMPLDLSNTSHQIMDTPDEEFPSPNLLRACS 77
 DB 176 PPTSSQYLPPLPAQSPSPSPSYLAPTSSYSSSSSSSYLSHSQSYLPS----- 228
 QY 78 GPQOASSGPPNNHSSOSTL---RPPLPPPHNHTLSHHSSANSILNRSLNRRSQHAPA 134
 DB 229 ---QASPSRFSQGHKSQQLLAPPSPAPPNGKTVQOEPL-----ESGANKSTEQQAA 280
 QY 135 PAPNDLATTPESVQLQDSWLVNSVPLETRHFLFKTSSGSTPLFSSSP-----GYPL 187
 DB 281 PEPDPSTMTQ---EQOQYWRQHLLSLQORTKVLPHGKGVVAKDTPPEVKEEVTVA 338
 QY 188 TSGTVTP---PPRLPNT 203
 DB 339 TSQVPSPSPSEEPPLPPN 357

RESULT 3

ABI3_ARATH STANDARD; PRT; 720 AA.
 AC Q01593;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Abscissic acid-insensitive protein 3.
 GN ABI3 OR AT3G24650 OR MSD24.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Fruit;
 RX MEDLINE=93076109; PubMed=1359917;
 RA Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F.,
 RA Goodman H.M.;
 RT "Isolation of the Arabidopsis ABI3 gene by positional cloning."
 RL Plant Cell 4:1251-1261(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20363039; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 PL, TAC
 RT and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC -1- FUNCTION: COULD PARTICIPATE IN ABSICISIC ACID-REGULATED GENE
 CC EXPRESSION DURING SEED DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: Contains 1 TF-B3 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X68141; CAA48241.1; -.
 DR EMBL: AP000740; BAB01214.1; -.
 DR PIR: JQ1676; JQ1676.
 DR InterPro: IPR003340; TF_B3.
 DR Pfam: PF02362; B3; 1.
 KW Developmental protein; Transcription regulation; DNA-binding;
 KW Activator; Nuclear protein. TF-B3.
 FT DOMAIN 532 720
 SQ SEQUENCE 720 AA; 79500 MW; D6A0F740D733060F CRC64;

Query Match 6.3%; Score 126.5; DB 1; Length 720;
 Best Local Similarity 21.6%; Pred. No. 0.21;
 Matches 77; Conservative 40; Mismatches 129; Indels 111; Gaps 17;
 QY 20 PPTSSPLSLQAQLPSSNNPPVSCMPLLDLSNTSHQIMDNPDEEF----- 66
 Db 378 PP--QPEFLPLESPSPPPPPQSGMP-----HQQFPMPPTSOYNQFGDPTGFNGYN 428
 QY 67 SPNSY-----LLRACSGPQQAASSGPPNHHQSQTLRPLPPPHNHTLSHHHS 113
 Db 429 MNPYQYVPVAGQMRDQLRLLCSSATKEA-----RKRMARQRRLSHHHR 475
 QY 114 SANSLNRLNLRNRRIQH-----APAPAPNDLATTPEVQLQDSVWLNNSVNPTRHFL 167
 Db 476 HNNN-NNNNNNNQQTGETCAVAPQLNPVATATG-----GTWTFWNPVAPVPPQL- 529
 QY 168 FKTSSGTPLFSSSPGYPVLSGVYTPPRLPNTFSRKAFKLKPKSKYCSWKCAALS 227
 Db 530 -----PPVMTQLPTMDRAGSASAMPRQVVPDR---RQGWKPEKRLFLQLVKQS 579
 QY 228 ATAAALLAILLAYIVPWSLNKSSIDSGEAEVGRVTVQVPPGV-----FW---R 275
 Db 580 DVGN-----LGRVLPRKAEATHLEARDGSLAME-DIGTSRVNMYRYFWPNKK 631
 QY 276 SQIHT--SOPOPKFN-----ISLGDALFGVYIRRL-----PPSHA 311
 Db 632 SRMYLLENTGDFVKTNGLEGDFVIYSDVKCGKYLIRGVKVPQSGQKPEAPPSSA 688

RESULT 4
 EXTN_TOBAC STANDARD; PRT; 620 AA.
 ID AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
 GN HRGPNT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxId=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Xanthi; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 MAIN ROOT.
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 GLYCOSYLATED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X13885; CAA32090.1; -.
 DR PIR: S06733; S06733.
 DR InterPro: IPR000480; Glutelin.
 DR InterPro: IPR002965; P-Rich_extensin.
 DR PRINTS: PR00211; GLUTELIN.
 DR PRINTS: PR01217; PRICHEXTENS.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 6.2%; Score 126; DB 1; Length 620;
 Best Local Similarity 24.1%; Pred. No. 0.19;
 Matches 59; Conservative 17; Mismatches 93; Indels 76; Gaps 11;
 QY 3 HAANKGRKPSAAGRPIPTSPSLPSAQLPSSH-----PPVSCMPLLDLS 51
 Db 133 HLPSSGQRPSPSHGHAPP-SGGHTPPRGQHPSPSRHSGHPPTTAQP----- 187
 QY 52 NTSHOIMTNDDEEPSNYSLLRACGPO---QASSGPPNHHSGSTLRP-----PLPP 102
 Db 188 -----PTPIYSPS-----PQVQPPPTYSPPPTTHVQTPSPSRGHQPQPP 228
 QY 103 PHNHT-LSHHSSANSNLSLTNRSSQIHAPAPAPNDLATTPEVQLQDSVWLNNSVPL 161
 Db 229 THRAPPHTRHA-----PTHQPSLRLHLPSPRRQPPPTYSPPPP 270
 QY 162 ETRHFLKTSSTPLFSSSPGY-PLTSGTVYTPPRL-----LPRNTFSRKAFKLK 213
 Db 271 -----AVAQSPQSPSTYSPPPTYSPPPPSPPIYSPPPPTPTPTPTSPPPPAYV 325
 QY 214 KPSKY 218
 Db 326 PPPTY 330

RESULT 5
 WS14_HUMAN STANDARD; PRT; 852 AA.
 ID WS14_HUMAN
 AC Q9NF71; Q96E48; Q9BY03; Q9BY04; Q9BY05; Q9BY06; Q9Y2P3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Williams-Beuren syndrome chromosome region 14 protein (WS basic-helix-
 loop-helix leucine zipper protein) (WS-BHLH) (Mix interactor).
 GN WBSCL14 OR MIO.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20241700; PubMed=10780788;
 RA de Luis O., Valero M.C., Perez Jurado L.A.;
 RT "WBSCL14, a putative transcription factor gene deleted in Williams-
 Beuren syndrome: complete characterisation of the human gene and the


```
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
KW Nuclear protein; Zinc_finger; Metal-binding; Repeat.
FT ZN_FING 346 368 C2H2-TYPE 1.
FT ZN_FING 374 397 C2H2-TYPE 2.
FT DOMAIN 8 35 POLY-GLN.
FT DOMAIN 98 104 POLY-ASN.
FT DOMAIN 143 157 POLY-ASN.
FT DOMAIN 173 177 POLY-ALA.
FT DOMAIN 240 245 POLY-HIS.
FT DOMAIN 417 420 POLY-SER.
FT DOMAIN 421 433 POLY-ASN.
FT DOMAIN 441 450 POLY-ALA.
SQ SEQUENCE 490 AA; 54382 MW; 4D0DA8DE43F171ED CRC64;

Query Match
Best Local Similarity 6.1%; Score 123; DB 1; Length 490;
Matches 62; Conservative 38; Mismatches 95; Indels 54; Gaps 15;

QY 22 TSSPSPLLPSAQLPSHNPVPCQPLLDSTNSHQIMTDNDEEFPNSYL-----LRA 75
DQ 164 TAAANNNNAAAAAYSPFTAMPTPOD-----QQYM-----FNPASYISHYSAYNS 212
QY 76 CSGPOQASSGPPNHHSSOTLRPLPPPHNTLSHHSSA-NSLNRSLSLNNRSHIAPA 134
DQ 213 NNNGNNAANG-SNNSSASAPAPCPPHH---HHSTNNLNGGAVTN---NAPQ 265
QY 135 PAPNDLATTPEVQLQD---SWVLSNV-PLETRH---FLFKTSSGSTPLFSSSSPGYPL 187
DQ 266 HHPT-IITDQFOFQQLQPNLNLINPAQPLHLPCKGKINTMPQPRTPAPNHPAPV 324
QY 188 TS-----GVYVTPP-----RLPRNFTSKAKFLKPKSYCSWKCAALSAIAALLAIL 239
DQ 325 PSSNPVSNLVPAPSDHKYIHQCFCEKFKR-----SW-----LKRHLSSHQ 371
QY 240 AVFVPMWL 248
DQ 372 RHFLCPWCL 380

RESULT 9
ID BBCL_YEAST STANDARD; PRT: 1157 AA.
AC P47068; P47067; Q8X1F4;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin tail region-interacting protein Wt1l (Bbcl1 protein).
GN BBCL1 OR Wt1l OR YTL020C/YJL021C OR J1305/J1286.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA To Van D., Perca J., Jacq C.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 699-765 FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR
RP LOCATION.
RX MEDLINE=21898311; PubMed=11901111;
RA Mochida J., Yamamoto T., Fujimura-Kanada K., Tanaka K.;
RT The novel adaptor protein, Mt1lp, and Vr1lp, a homolog of
RT Wiskott-Aldrich syndrome protein-interacting protein (WIP), may
RT antagonistically regulate type I myosins in Saccharomycetes
RT cerevisiae.
RL Genetics 160:923-934(2002).
RN [3]
RP SEQUENCE OF 775-1157 FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in the regulation of actin cytoskeleton.
CC -!- SUBUNIT: Binds to the SH3 domains of the type I myosins MYO3 and
```

```
CC MYO5.
CC -!- SUBCELLULAR LOCATION: Co-localizes with cortical actin patches.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 729 and 732.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; Z49295; CAA89311.1; ALT_FRAME.
DR EMBL; Z49296; CAA89312.1; ALT_INIT.
DR EMBL; AF373805; AAL57239.1; -.
DR SGD; S0003557; BBL1
DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.
DR GO; GO:0017024; F:myosin I binding activity; IPI.
DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; IGI.
DR GO; GO:0007011; P:regulation of cytoskeleton; IGI.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain; Coiled coil.
FT DOMAIN 5 69 SH3.
FT DOMAIN 234 301 COILED COIL (POTENTIAL).
FT DOMAIN 278 412 GLU-RICH.
FT DOMAIN 356 430 COILED COIL (POTENTIAL).
FT DOMAIN 674 830 PRO-RICH.
SQ SEQUENCE 1157 AA; 128296 MW; 88A5899B89CCE8ED CRC64;

Query Match
Best Local Similarity 6.1%; Score 123; DB 1; Length 1157;
Matches 94; Conservative 57; Mismatches 154; Indels 186; Gaps 21;

QY 11 PSAAEGRIPPTSPSPLLPSAQLPSHNPVPCQPLLDSTNSHQIMTDNDEEFPNS 70
DQ 707 PVPSPAPPVP--APPALSAPSPVPVPVPPVSSAPPALSAPSPVPPTPPAPPAPAP 764
QY 71 YLLRACSGPQQASSSGPP-----NHSQSTLRPLP-----PP-----HNHT 107
DQ 765 LALPKHNEVEHVKSSAPLPVSEYHPMPNTAPLPAPPVPPATFEEDSEPTATHSHT 824
QY 108 L-----SHHH-----SSANSLRN-----SLTNRRSIIHAPAPAP 137
DQ 825 APSPPPHQNVTAFTSPSMSTOQVRPTSVLSGAEKESRTLPHPVPSLTNR-----PVDSF 878
QY 138 NDLATTPESVOL-----QDSWVLSNVP-----QDSWVLSNVP-----160
DQ 879 HESDTPKVASIRRTTHDVGESISNNVKIEFNAQERWINKSAPSAISNLKLNFLMEIDD 938
QY 161 -----LETRHFLFTSS---GSTPLFSSSSPGYPLTS-----GTVYTPPRL 200
DQ 939 HFISKRLHQKVVVRFYELFENYSQLRFSLTENSTSPKTVTLQERFSPVETQSARIL 998
QY 201 PRNTERKAFKLLKPKSKYCS-----WKCAALSAIAAALLAILLAYF 242
DQ 999 DEYA---QRFNKAVVEKSHSLNSHIGAKNFVQSIVSEFKDEVIOPIGARTFGATLSY- 1054
QY 243 IVPWSLKNSSIDSGAEVQ-----RRVTEVPPGPFVRSQIHLSQPFQKFNLSLGLKDALF 298
DQ 1055 -----KPEEGIEQLMSLQKIKPG-----DILVIRKAKFAHKKIKNEII 1095
QY 299 GVIIRGLPPSH--AQYDFMERLDGKEKWSVSPRRRSIOTLVQNEAVFYQYLDVGLW 356
DQ 1096 NVGMDSAAPYSVVTDYDFT-----KNKFRVIENHEGK-----IIONS-----Y 1134
QY 357 HIAFYNDGDKD 367
DQ 1134
```

```
Db 1135 KLSHMSGKLIK 1145

RESULT 10
GSRL_HUMAN
ID GSRL_HUMAN STANDARD; PRT; 1509 AA.
AC GSNZM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma tumor suppressor candidate region gene 1 protein.
GN GLTSCR1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT region";
RL Genomics 64:44-50(2000).
CC -1- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC placenta, skeletal muscle, and pancreas, and at lower levels in
CC lung, liver, and kidney.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF182077; AAF62874.1;
DR Genew; HGNC:4332; GLTSCR1.
DR MIM; 605690;
DR DOMAIN 37 45 POLY-GLY.
FT DOMAIN 884 889 POLY-PRO.
FT DOMAIN 1214 1225 POLY-SER.
FT DOMAIN 1282 1286 POLY-PRO.
FT DOMAIN 1294 1304 POLY-PRO.
SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443C6B821 CRC64;

Query Match 6.0%; Score 121; DB 1; Length 1509;
Best Local Similarity 23.3%; Pred. No. 1.2;
Matches 62; Conservative 26; Mismatches 94; Indels 84; Gaps 12;

QY 11 PSAEAGRIPPTSSPSLLPSAQL-PSHNPPVPVSCOMPLLDNSTSHQIMDTPDE---EF 66
Db 801 PTA-PGPPQPLRQSQPEGPLPPAPHLPPSTSSAVASSSTSRRLPAPTSPDQLQF 859
QY 67 SPNSYLRLACSGPQQAASSGPPNHHQSQTLRPLPPPHNH-----TLSSH 111
Db 860 PPS-----QGPKSPPTPTLHLVPEAAPPPPPPTFMVTTFPALPQPKALERF 912
QY 112 H-----SSANSLNLSLNRRSQI-----HAPAPAPND 139
Db 913 HQVPSGIILONKAGAPAPQSTSLGLTSPNAASVLVSGQAPSGPTTAPSHAPAPAP-- 970
QY 140 LATTPEVSQLODSWVLSNVPLETRHFLFKTSGSTPLFSSSPGYPPLTSGTVY----- 193
Db 971 MAATG-----LPPLPAENKAFASNLPLNVAKAASSGPGRP--SGLOYESKLSG 1018
QY 194 -TPPPRLPNTFSKAFKLKPKSKY 218
Db 1019 LKKPPTLPQ-----SKACFLHLHKKH 1040

RESULT 11

MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RP TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -1- SIMILARITY: Contains 2 VWFC domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L21998; AAB95295.1;
DR EMBL; M74027; AAB95875.1;
DR EMBL; M94131; AAB9163.1;
DR EMBL; M94132; AAB9164.1;
DR PIR; A49963; A43932.
DR Genew; HGNC:7512; MUC2.
DR MIM; 158370;
DR GO; GO:0005803; C:secretory vesicle; TAS.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR006209; EGF_like.
```

```

DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR001007; VWF_C.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF00007; Cys_knot; 1.
DR Pfam: PF01826; TIL; 1.
DR Pfam: PF00093; vvc; 1.
DR Pfam: PF00094; vvd; 4.
DR SMART: SM00214; vvc; 2.
DR SMART: SM00216; vvd; 4.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01208; VWF_C_1; 2.
DR PROSITE: PS0184; VWF_C_2; 2.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 20
FT CHAIN 21 5179
FT DOMAIN 1401 1747
FT REPEAT 1401 1416
FT REPEAT 1417 1432
FT REPEAT 1433 1448
FT REPEAT 1449 1464
FT REPEAT 1465 1471
FT REPEAT 1472 1478
FT REPEAT 1479 1494
FT REPEAT 1495 1517
FT REPEAT 1518 1533
FT REPEAT 1534 1556
FT REPEAT 1557 1572
FT REPEAT 1573 1596
FT REPEAT 1597 1612
FT REPEAT 1613 1635
FT REPEAT 1636 1651
FT REPEAT 1652 1675
FT REPEAT 1676 1683
FT REPEAT 1684 1699
FT REPEAT 1700 1715
FT REPEAT 1716 1731
FT REPEAT 1732 1747
FT REPEAT 1748 1763
FT REPEAT 1764 1786
FT REPEAT 1787 1802
FT REPEAT 1803 1820
FT REPEAT 1821 1839
FT REPEAT 1840 1857
FT REPEAT 1858 1875
FT REPEAT 1876 1893
FT REPEAT 1894 1911
FT REPEAT 1912 1929
FT REPEAT 1930 1947
FT REPEAT 1948 1965
FT REPEAT 1966 1983
FT REPEAT 1984 2001
FT REPEAT 2002 2019
FT REPEAT 2020 2037
FT REPEAT 2038 2055
FT REPEAT 2056 2073
FT REPEAT 2074 2091
FT REPEAT 2092 2109
FT REPEAT 2110 2127
FT REPEAT 2128 2145
FT REPEAT 2146 2163
FT REPEAT 2164 2181
FT REPEAT 2182 2199
FT REPEAT 2200 2217
FT REPEAT 2218 2235
FT REPEAT 2236 2253
FT REPEAT 2254 2271
FT REPEAT 2272 2289
FT REPEAT 2290 2307
FT REPEAT 2308 2325
FT REPEAT 2326 2343
FT REPEAT 2344 2361
FT REPEAT 2362 2379
FT REPEAT 2380 2397
FT REPEAT 2398 2415
FT REPEAT 2416 2433
FT REPEAT 2434 2451
FT REPEAT 2452 2469
FT REPEAT 2470 2487
FT REPEAT 2488 2505
FT REPEAT 2506 2523
FT REPEAT 2524 2541
FT REPEAT 2542 2559
FT REPEAT 2560 2577
FT REPEAT 2578 2595
FT REPEAT 2596 2613
FT REPEAT 2614 2631
FT REPEAT 2632 2649
FT REPEAT 2650 2667
FT REPEAT 2668 2685
FT REPEAT 2686 2703
FT REPEAT 2704 2721
FT REPEAT 2722 2739
FT REPEAT 2740 2757
FT REPEAT 2758 2775
FT REPEAT 2776 2793
FT REPEAT 2794 2811
FT REPEAT 2812 2829
FT REPEAT 2830 2847
FT REPEAT 2848 2865
FT REPEAT 2866 2883
FT REPEAT 2884 2901
FT REPEAT 2902 2919
FT REPEAT 2920 2937
FT REPEAT 2938 2955
FT REPEAT 2956 2973
FT REPEAT 2974 2991
FT REPEAT 2992 3009
FT REPEAT 3010 3027
FT REPEAT 3028 3045
FT REPEAT 3046 3063
FT REPEAT 3064 3081
FT REPEAT 3082 3099
FT REPEAT 3100 3117
FT REPEAT 3118 3135
FT REPEAT 3136 3153
FT REPEAT 3154 3171
FT REPEAT 3172 3189
FT REPEAT 3190 3207
FT REPEAT 3208 3225
FT REPEAT 3226 3243
FT REPEAT 3244 3261
FT REPEAT 3262 3279
FT REPEAT 3280 3297
FT REPEAT 3298 3315
FT REPEAT 3316 3333
FT REPEAT 3334 3351
FT REPEAT 3352 3369
FT REPEAT 3370 3387
FT REPEAT 3388 3405
FT REPEAT 3406 3423
FT REPEAT 3424 3441
FT REPEAT 3442 3459
FT REPEAT 3460 3477
FT REPEAT 3478 3495
FT REPEAT 3496 3513
FT REPEAT 3514 3531
FT REPEAT 3532 3549
FT REPEAT 3550 3567
FT REPEAT 3568 3585
FT REPEAT 3586 3603
FT REPEAT 3604 3621
FT REPEAT 3622 3639
FT REPEAT 3640 3657
FT REPEAT 3658 3675
FT REPEAT 3676 3693
FT REPEAT 3694 3711
FT REPEAT 3712 3729
FT REPEAT 3730 3747
FT REPEAT 3748 3765
FT REPEAT 3766 3783
FT REPEAT 3784 3801
FT REPEAT 3802 3819
FT REPEAT 3820 3837
FT REPEAT 3838 3855
FT REPEAT 3856 3873
FT REPEAT 3874 3891
FT REPEAT 3892 3909
FT REPEAT 3910 3927
FT REPEAT 3928 3945
FT REPEAT 3946 3963
FT REPEAT 3964 3981
FT REPEAT 3982 3999
FT REPEAT 4000 4017
FT REPEAT 4018 4035
FT REPEAT 4036 4053
FT REPEAT 4054 4071
FT REPEAT 4072 4089
FT REPEAT 4090 4107
FT REPEAT 4108 4125
FT REPEAT 4126 4143
FT REPEAT 4144 4161
FT REPEAT 4162 4179
FT REPEAT 4180 4197
FT REPEAT 4198 4215
FT REPEAT 4216 4233
FT REPEAT 4234 4251
FT REPEAT 4252 4269
FT REPEAT 4270 4287
FT REPEAT 4288 4305
FT REPEAT 4306 4323
FT REPEAT 4324 4341
FT REPEAT 4342 4359
FT REPEAT 4360 4377
FT REPEAT 4378 4395
FT REPEAT 4396 4413
FT REPEAT 4414 4431
FT REPEAT 4432 4449
FT REPEAT 4450 4467
FT REPEAT 4468 4485
FT REPEAT 4486 4503
FT REPEAT 4504 4521
FT REPEAT 4522 4539
FT REPEAT 4540 4557
FT REPEAT 4558 4575
FT REPEAT 4576 4593
FT REPEAT 4594 4611
FT REPEAT 4612 4629
FT REPEAT 4630 4647
FT REPEAT 4648 4665
FT REPEAT 4666 4683
FT REPEAT 4684 4701
FT REPEAT 4702 4719
FT REPEAT 4720 4737
FT REPEAT 4738 4755
FT REPEAT 4756 4773
FT REPEAT 4774 4791
FT REPEAT 4792 4809
FT REPEAT 4810 4827
FT REPEAT 4828 4845
FT REPEAT 4846 4863
FT REPEAT 4864 4881
FT REPEAT 4882 4899
FT REPEAT 4900 4917
FT REPEAT 4918 4935
FT REPEAT 4936 4953
FT REPEAT 4954 4971
FT REPEAT 4972 4989
FT REPEAT 4990 5007
FT REPEAT 5008 5025
FT REPEAT 5026 5043
FT REPEAT 5044 5061
FT REPEAT 5062 5079
FT REPEAT 5080 5097
FT REPEAT 5098 5115
FT REPEAT 5116 5133
FT REPEAT 5134 5151
FT REPEAT 5152 5169
FT REPEAT 5170 5187
FT REPEAT 5188 5205
FT REPEAT 5206 5223
FT REPEAT 5224 5241
FT REPEAT 5242 5259
FT REPEAT 5260 5277
FT REPEAT 5278 5295
FT REPEAT 5296 5313
FT REPEAT 5314 5331
FT REPEAT 5332 5349
FT REPEAT 5350 5367
FT REPEAT 5368 5385
FT REPEAT 5386 5403
FT REPEAT 5404 5421
FT REPEAT 5422 5439
FT REPEAT 5440 5457
FT REPEAT 5458 5475
FT REPEAT 5476 5493
FT REPEAT 5494 5511
FT REPEAT 5512 5529
FT REPEAT 5530 5547
FT REPEAT 5548 5565
FT REPEAT 5566 5583
FT REPEAT 5584 5601
FT REPEAT 5602 5619
FT REPEAT 5620 5637
FT REPEAT 5638 5655
FT REPEAT 5656 5673
FT REPEAT 5674 5691
FT REPEAT 5692 5709
FT REPEAT 5710 5727
FT REPEAT 5728 5745
FT REPEAT 5746 5763
FT REPEAT 5764 5781
FT REPEAT 5782 5799
FT REPEAT 5800 5817
FT REPEAT 5818 5835
FT REPEAT 5836 5853
FT REPEAT 5854 5871
FT REPEAT 5872 5889
FT REPEAT 5890 5907
FT REPEAT 5908 5925
FT REPEAT 5926 5943
FT REPEAT 5944 5961
FT REPEAT 5962 5979
FT REPEAT 5980 5997
FT REPEAT 5998 6015
FT REPEAT 6016 6033
FT REPEAT 6034 6051
FT REPEAT 6052 6069
FT REPEAT 6070 6087
FT REPEAT 6088 6105
FT REPEAT 6106 6123
FT REPEAT 6124 6141
FT REPEAT 6142 6159
FT REPEAT 6160 6177
FT REPEAT 6178 6195
FT REPEAT 6196 6213
FT REPEAT 6214 6231
FT REPEAT 6232 6249
FT REPEAT 6250 6267
FT REPEAT 6268 6285
FT REPEAT 6286 6303
FT REPEAT 6304 6321
FT REPEAT 6322 6339
FT REPEAT 6340 6357
FT REPEAT 6358 6375
FT REPEAT 6376 6393
FT REPEAT 6394 6411
FT REPEAT 6412 6429
FT REPEAT 6430 6447
FT REPEAT 6448 6465
FT REPEAT 6466 6483
FT REPEAT 6484 6501
FT REPEAT 6502 6519
FT REPEAT 6520 6537
FT REPEAT 6538 6555
FT REPEAT 6556 6573
FT REPEAT 6574 6591
FT REPEAT 6592 6609
FT REPEAT 6610 6627
FT REPEAT 6628 6645
FT REPEAT 6646 6663
FT REPEAT 6664 6681
FT REPEAT 6682 6699
FT REPEAT 6700 6717
FT REPEAT 6718 6735
FT REPEAT 6736 6753
FT REPEAT 6754 6771
FT REPEAT 6772 6789
FT REPEAT 6790 6807
FT REPEAT 6808 6825
FT REPEAT 6826 6843
FT REPEAT 6844 6861
FT REPEAT 6862 6879
FT REPEAT 6880 6897
FT REPEAT 6898 6915
FT REPEAT 6916 6933
FT REPEAT 6934 6951
FT REPEAT 6952 6969
FT REPEAT 6970 6987
FT REPEAT 6988 7005
FT REPEAT 7006 7023
FT REPEAT 7024 7041
FT REPEAT 7042 7059
FT REPEAT 7060 7077
FT REPEAT 7078 7095
FT REPEAT 7096 7113
FT REPEAT 7114 7131
FT REPEAT 7132 7149
FT REPEAT 7150 7167
FT REPEAT 7168 7185
FT REPEAT 7186 7203
FT REPEAT 7204 7221
FT REPEAT 7222 7239
FT REPEAT 7240 7257
FT REPEAT 7258 7275
FT REPEAT 7276 7293
FT REPEAT 7294 7311
FT REPEAT 7312 7329
FT REPEAT 7330 7347
FT REPEAT 7348 7365
FT REPEAT 7366 7383
FT REPEAT 7384 7401
FT REPEAT 7402 7419
FT REPEAT 7420 7437
FT REPEAT 7438 7455
FT REPEAT 7456 7473
FT REPEAT 7474 7491
FT REPEAT 7492 7509
FT REPEAT 7510 7527
FT REPEAT 7528 7545
FT REPEAT 7546 7563
FT REPEAT 7564 7581
FT REPEAT 7582 7599
FT REPEAT 7600 7617
FT REPEAT 7618 7635
FT REPEAT 7636 7653
FT REPEAT 7654 7671
FT REPEAT 7672 7689
FT REPEAT 7690 7707
FT REPEAT 7708 7725
FT REPEAT 7726 7743
FT REPEAT 7744 7761
FT REPEAT 7762 7779
FT REPEAT 7780 7797
FT REPEAT 7798 7815
FT REPEAT 7816 7833
FT REPEAT 7834 7851
FT REPEAT 7852 7869
FT REPEAT 7870 7887
FT REPEAT 7888 7905
FT REPEAT 7906 7923
FT REPEAT 7924 7941
FT REPEAT 7942 7959
FT REPEAT 7960 7977
FT REPEAT 7978 7995
FT REPEAT 7996 8013
FT REPEAT 8014 8031
FT REPEAT 8032 8049
FT REPEAT 8050 8067
FT REPEAT 8068 8085
FT REPEAT 8086 8103
FT REPEAT 8104 8121
FT REPEAT 8122 8139
FT REPEAT 8140 8157
FT REPEAT 8158 8175
FT REPEAT 8176 8193
FT REPEAT 8194 8211
FT REPEAT 8212 8229
FT REPEAT 8230 8247
FT REPEAT 8248 8265
FT REPEAT 8266 8283
FT REPEAT 8284 8301
FT REPEAT 8302 8319
FT REPEAT 8320 8337
FT REPEAT 8338 8355
FT REPEAT 8356 8373
FT REPEAT 8374 8391
FT REPEAT 8392 8409
FT REPEAT 8410 8427
FT REPEAT 8428 8445
FT REPEAT 8446 8463
FT REPEAT 8464 8481
FT REPEAT 8482 8499
FT REPEAT 8500 8517
FT REPEAT 8518 8535
FT REPEAT 8536 8553
FT REPEAT 8554 8571
FT REPEAT 8572 8589
FT REPEAT 8590 8607
FT REPEAT 8608 8625
FT REPEAT 8626 8643
FT REPEAT 8644 8661
FT REPEAT 8662 8679
FT REPEAT 8680 8697
FT REPEAT 8698 8715
FT REPEAT 8716 8733
FT REPEAT 8734 8751
FT REPEAT 8752 8769
FT REPEAT 8770 8787
FT REPEAT 8788 8805
FT REPEAT 8806 8823
FT REPEAT 8824 8841
FT REPEAT 8842 8859
FT REPEAT 8860 8877
FT REPEAT 8878 8895
FT REPEAT 8896 8913
FT REPEAT 8914 8931
FT REPEAT 8932 8949
FT REPEAT 8950 8967
FT REPEAT 8968 8985
FT REPEAT 8986 9003
FT REPEAT 9004 9021
FT REPEAT 9022 9039
FT REPEAT 9040 9057
FT REPEAT 9058 9075
FT REPEAT 9076 9093
FT REPEAT 9094 9111
FT REPEAT 9112 9129
FT REPEAT 9130 9147
FT REPEAT 9148 9165
FT REPEAT 9166 9183
FT REPEAT 9184 9201
FT REPEAT 9202 9219
FT REPEAT 9220 9237
FT REPEAT 9238 9255
FT REPEAT 9256 9273
FT REPEAT 9274 9291
FT REPEAT 9292 9309
FT REPEAT 9310 9327
FT REPEAT 9328 9345
FT REPEAT 9346 9363
FT REPEAT 9364 9381
FT REPEAT 9382 9399
FT REPEAT 9400 9417
FT REPEAT 9418 9435
FT REPEAT 9436 9453
FT REPEAT 9454 9471
FT REPEAT 9472 9489
FT REPEAT 9490 9507
FT REPEAT 9508 9525
FT REPEAT 9526 9543
FT REPEAT 9544 9561
FT REPEAT 9562 9579
FT REPEAT 9580 9597
FT REPEAT 9598 9615
FT REPEAT 9616 9633
FT REPEAT 9634 9651
FT REPEAT 9652 9669
FT REPEAT 9670 9687
FT REPEAT 9688 9705
FT REPEAT 9706 9723
FT REPEAT 9724 9741
FT REPEAT 9742 9759
FT REPEAT 9760 9777
FT REPEAT 9778 9795
FT REPEAT 9796 9813
FT REPEAT 9814 9831
FT REPEAT 9832 9849
FT REPEAT 9850 9867
FT REPEAT 9868 9885
FT REPEAT 9886 9903
FT REPEAT 9904 9921
FT REPEAT 9922 9939
FT REPEAT 9940 9957
FT REPEAT 9958 9975
FT REPEAT 9976 9993
FT REPEAT 9994 10011
FT REPEAT 10012 10030
FT REPEAT 10031 10048
FT REPEAT 10049 10066
FT REPEAT 10067 10084
FT REPEAT 10085 10102
FT REPEAT 10103 10120
FT REPEAT 10121 10138
FT REPEAT 10139 10156
FT REPEAT 10157 10173
FT REPEAT 10174 10191
FT REPEAT 10192 10209
FT REPEAT 10210 10227
FT REPEAT 10228 10245
FT REPEAT 10246 10263
FT REPEAT 10264 10281
FT REPEAT 10282 10299
FT REPEAT 10300 10317
FT REPEAT 10318 10335
FT REPEAT 10336 10353
FT REPEAT 10354 10371
FT REPEAT 10372 10389
FT REPEAT 10390 10407
FT REPEAT 10408 10425
FT REPEAT 10426 10443
FT REPEAT 10444 10461
FT REPEAT 10462 10479
FT REPEAT 10480 10497
FT REPEAT 10498 10515
FT REPEAT 10516 10533
FT REPEAT 10534 10551
FT REPEAT 10552 10569
FT REPEAT 10570 10587
FT REPEAT 10588 10605
FT REPEAT 10606 10623
FT REPEAT 10624 10641
FT REPEAT 10642 10659
FT REPEAT 10660 10677
FT REPEAT 10678 10695
FT REPEAT 10696 10713
FT REPEAT 10714 10731
FT REPEAT 10732 10749
FT REPEAT 10750 10767
FT REPEAT 10768 10785
FT REPEAT 10786 10803
FT REPEAT 10804 10821
FT REPEAT 10822 10839
FT REPEAT 10840 10857
FT REPEAT 10858 10875
FT REPEAT 10876 10893
FT REPEAT 10894 10911
FT REPEAT 10912 10929
FT REPEAT 10930 10947
FT REPEAT 10948 10965
FT REPEAT 10966 10983
FT REPEAT 10984 11001
FT REPEAT 11002 11019
FT REPEAT 11020 11037
FT REPEAT 11038 11055
FT REPEAT 11056 11073
FT REPEAT 11074 11091
FT REPEAT 11092 11109
FT REPEAT 11110 11127
FT REPEAT 11128 11145
FT REPEAT 11146 11163
FT REPEAT 11164 11181
FT REPEAT 11182 11199
FT REPEAT 11200 11217
FT REPEAT 11218 11235
FT REPEAT 11236 11253
FT REPEAT 11254 11271
FT REPEAT 11272 11289
FT REPEAT 11290 11307
FT REPEAT 11308 11325
FT REPEAT 11326 11343
FT REPEAT 11344 11361
FT REPEAT 11362 11379
FT REPEAT 11380 11397
FT REPEAT 11398 11415
FT REPEAT 11416 11433
FT REPEAT 11434 11451
FT REPEAT 11452 11469
FT REPEAT 11470 11487
FT REPEAT 11488 11505
FT REPEAT 11506 11523
FT REPEAT 11524 11541
FT REPEAT 11542 11559
FT REPEAT 11560 11577
FT REPEAT 11578 11595
FT REPEAT 11596 11613
FT REPEAT 11614 11631
FT REPEAT 11632 11649
FT REPEAT 11650 11667
FT REPEAT 11668 11685
FT REPEAT 11686 11703
FT REPEAT 11704 11721
FT REPEAT 11722 11739
FT REPEAT 11740 11757
FT REPEAT 11758 11775
FT REPEAT 11776 11793
FT REPEAT 11794 11811
FT REPEAT 11812 11829
FT REPEAT 11830 11847
FT REPEAT 11848 11865
FT REPEAT 11866 11883
FT REPEAT 11884 11901
FT REPEAT 11902 11919
FT REPEAT 11920 11937
FT REPEAT 11938 11955
FT REPEAT 11956 11973
FT REPEAT 11974 11991
FT REPEAT 11992 12009
FT REPEAT 12010 12027
FT REPEAT 12028 12045
FT REPEAT 12046 12063
FT REPEAT 12064 12081
FT REPEAT 12082 12099
FT REPEAT 12100 12117
FT REPEAT 12118 12135
FT REPEAT 12136 12153
FT REPEAT 12154 12171
FT REPEAT 12172 12189
FT REPEAT 12190 12207
FT REPEAT 12208 12225
FT REPEAT 12226 12243
FT REPEAT 12244 12261
FT REPEAT 12262 12279
FT REPEAT 12280 12297
FT REPEAT 12298 12315
FT REPEAT 12316 12333
FT REPEAT 12334 12351
FT REPEAT 12352 12369
FT REPEAT 12370 12387
FT REPEAT 12388 12405
FT REPEAT 12406 12423
FT REPEAT 12424 12441
FT REPEAT 12442 12459
FT REPEAT 12460 12477
FT REPEAT 12478 12495
FT REPEAT 12496 12513
FT REPEAT 12514 12531
FT REPEAT 12532 12549
FT REPEAT 12550 12567
FT REPEAT 12568 12585
FT REPEAT 12586 12603
FT REPEAT 12604 12621
FT REPEAT 12622 12639
FT REPEAT 12640 12657
FT REPEAT 12658 12675
FT REPEAT 12676 12693
FT REPEAT 12694 12711
FT REPEAT 12712 12729
FT REPEAT 12730 12747
FT REPEAT 12748 12765
FT REPEAT 12766 12783
FT REPEAT 12784 12801
FT REPEAT 12802 12819
FT REPEAT 12820 12837
FT REPEAT 12838 12855
FT REPEAT 12856 12873
FT REPEAT 12874 12891
FT REPEAT 12892 12909
FT REPEAT 12910 12927
FT REPEAT 12928 12945
FT REPEAT 12946 12963
FT REPEAT 12964 12981
FT REPEAT 12982 13000
FT REPEAT 13001 13018
FT REPEAT 13019 13036
FT REPEAT 13037 13054
FT REPEAT 13055 13072
FT REPEAT 13073 13090
FT REPEAT 13091 13108
FT REPEAT 13109 13126
FT REPEAT 13127 13144
FT REPEAT 13145 13162
FT REPEAT 13163 13180
FT REPEAT 13181 13198
FT REPEAT 13199 13216
FT REPEAT 13217 13234
FT REPEAT 13235 13252
FT REPEAT 13253 13269
FT REPEAT 13270 13287
FT REPEAT 13288 13305
FT REPEAT 13306 13323
FT REPEAT 13324 13341
FT REPEAT 13342 13359
FT REPEAT 13360 13377
FT REPEAT 13378 13395
FT REPEAT 13396 13413
FT REPEAT 13414 13431
FT REPEAT 13432 13449
FT REPEAT 13450 13467
FT REPEAT 13468 13485
FT REPEAT 13486 13503
FT REPEAT 13504 13521
FT REPEAT 13522 13539
FT REPEAT 13540 13557
FT REPEAT 13558 13575
FT REPEAT 13576 13593
FT REPEAT 13594 13611
FT REPEAT 13612 13629
FT REPEAT 13630 13647
FT REPEAT 13648 13665
FT REPEAT 13666 13683
FT REPEAT 13684 13701
FT REPEAT 13702 13719
FT REPEAT 13720 13737
FT REPEAT 13738 13755
FT REPEAT 13756 13773
FT REPEAT 13774 13791
FT REPEAT 13792 13809
FT REPEAT 13810 13827
FT REPEAT 13828 13845
FT REPEAT 13846 13863
FT REPEAT 13864 13881
FT REPEAT 13882 13899
FT REPEAT 13900 13917
FT REPEAT 13918 13935
FT REPEAT 13936 13953
FT REPEAT 13954 13971
FT REPEAT 13972 13989
FT REPEAT 13990 14007
FT REPEAT 14008 14025
FT REPEAT 14026 14043
FT REPEAT 14044 14061
FT REPEAT 14062 14079
FT REPEAT 14080 14097
FT REPEAT 14098 14115
FT REPEAT 14116 14133
FT REPEAT 14134 14151
FT REPEAT 14152 14169
FT REPEAT 14170 14187
FT REPEAT 14188 14205
FT REPEAT 14206 14223
FT REPEAT 14224 14241
FT REPEAT 14242 14259
FT REPEAT 14260 14277
FT REPEAT 14278 14295
FT REPEAT 14296 14313
FT REPEAT 14314 14331
FT REPEAT 14332 14349
FT REPEAT 14350 14367
FT REPEAT 14368 14385
FT REPEAT 14386 14403
FT REPEAT 14404 14421
FT REPEAT 14422 14439
FT REPEAT 14440 14457
FT REPEAT 14458 14475
FT REPEAT 14476 14493
FT REPEAT 14494 14511
FT REPEAT 14512 14529
FT REPEAT 14530 14547
FT REPEAT 14548 14565
FT REPEAT 14566 14583
FT REPEAT 14584 14601
FT REPEAT 14602 14619
FT REPEAT 14620 14637
FT REPEAT 14638 14655
FT REPEAT 14656 14673
FT REPEAT 14674 14691
FT REPEAT 14692 14709
FT REPEAT 14710 14727
FT REPEAT 14728 14745
FT REPEAT 14746 14763
FT REPEAT 14764 14781
FT REPEAT 14782 14799
FT REPEAT 14800 14817
FT REPEAT 14818 14835
FT REPEAT 14836 14853
FT REPEAT 14854 14871
FT REPEAT 14872 14889
FT REPEAT 14890 14907
FT REPEAT 14908 14925
FT REPEAT 14926 14943
FT REPEAT 14944 14961
FT REPEAT 14962 14979
FT REPEAT 14980 14997
FT REPEAT 14998 15015
FT REPEAT 15016 15033
FT REPEAT 15034 15051
FT REPEAT 15052 15069
FT REPEAT 15070 15087
FT REPEAT 15088 15105
FT REPEAT 15106 15123
FT REPEAT 15124 15141
FT REPEAT 15142 15159
FT REPEAT 15160 15177
FT REPEAT 15178 15195
FT REPEAT 15196 15213
FT REPEAT 15214 15231
FT REPEAT 15232 15249
FT REPEAT 15250 15267
FT REPEAT 15268 15285
FT REPEAT 15286 15303
FT REPEAT 15304 15321
FT REPEAT 15322 15339
FT REPEAT 15340 15357
FT REPEAT 15358 15375
FT REPEAT 15376 15393
FT REPEAT 15394 15411
FT REPEAT 15412 15429
FT REPEAT 15430 15447
FT REPEAT 15448 15465
FT REPEAT 15466 15483
FT REPEAT 15484 15501
FT REPEAT 15502 15519
FT REPEAT 15520 15537
FT REPEAT 15538 15555
FT REPEAT 15556 15573
FT REPEAT 15574 15591
FT REPEAT 15592 15609
FT REPEAT 15610 15627
FT REPEAT 15628 15645
FT REPEAT 15646 15663
FT REPEAT 15664 15681
FT REPEAT 15682 15699
FT REPEAT 15700 15717
FT REPEAT 15718 15735
FT REPEAT 15736 15753
FT REPEAT 15754 15771
FT REPEAT 15772 15789
FT REPEAT 15790 15807
FT REPEAT 15808 15825
FT REPEAT 15826 15843
FT REPEAT 15844 1586
```



```
QY 230 AAALLAILLAFYIPWLSKNSI--DSGEAE-----VGRRTQEVPP--GVF- 273
Db 308 -----AKVVVTRVYGGYVYEAQAQDEYDTPRHLLAPGSQDIYDPPVGRLLP 357
QY 274 --WRSGIHISOPQELKENSLSKDALFGVY-----IRGLPPS--HAQYD 314
Db 358 NOYQGEVDYTPMAVKG--PGRDPLLDYDVPVSVKGLPPSNHSHVYD 405

RESULT 14
V70K.EPMV
ID V70K.EPMV STANDARD; PRT; 649 AA.
AC 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 70 kDa protein.
OS Eggplant mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX NCBI_TaxID=12151;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021185; PubMed=2800336;
RA Osorio-Keese M.E., Keese P., Gibbs A.;
RT "Nucleotide sequence of the genome of eggplant mosaic tymovirus.";
RL Virology 172:547-554(1989).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04374; AAA43038.1; -
DR PIR; J00103; J00103.
DR InterPro; IPR004935; Tymo_45_70kDa.
DR Pfam; PF03251; Tymo_45kd_70kd; 1.
SQ SEQUENCE 649 AA; 70174 MW; 6639DFB7518AA80F CRC64;

Query Match 5.9%; Score 119.5; DB 1; Length 649;
Best Local Similarity 23.7%; Pred. No. 0.55;
Matches 72; Conservative 31; Mismatches 98; Indels 103; Gaps 17;

QY 3 HAANKGRKPSAAGRPPIPTSSPLPSAQLP-----SSH----- 37
Db 366 HSKRVGQSADLVQCQPPPP---PSMLPLLLLPGRGVKAPRLPTLASLSFGSHPIPYHTS 422
QY 38 NPPVSCQPLDLSNTSHQMDTNPDEEFSNLYLR--ACSGPQQAASSGPP----- 88
Db 423 SPPLIQFHQPPPPPSATFSV-----SPLGVLTTAFALNPTQSAERSCDPESFTPTL 473
QY 89 NIHSGSTLRPLPPPHNHTLSHHSSA-----NSLNRS-----LTNRRSQIHA- 132
Db 474 GHKTTLSRLPLPPPHSPHSQAORASALATVSNSETKNCPSPTVPPPLPN---HLHPL 530
QY 133 ----PAPANDLATTPESVQLQ---DSVNLN---VPLETRHFLFKTSGSPPLFSS 180
Db 531 LGETDPTTPRLQSPSSLSLRTFLDSAVISCDSSPVLPPSP-----SPSSHSSSSFS 585
QY 181 --SSGCVPL-----TSGVTYTPPPRLLP--RNFTFSRKAFKLK---PSKYC 219
Db 586 CTSPPRFPCYPPPPSALDILLFSTGCTDPPDPDPHPKHTIALERLRLQCATPFHSC 645
QY 220 SWKC 223
Db 646 DDVC 649
```

```
RESULT 15
M4K3_HUMAN
ID M4K3_HUMAN STANDARD; PRT; 894 AA.
AC Q81VH8; Q81VH7; Q9UDM5; Q9Y6R5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 3 (EC 2.7.1.37)
DE (MAPK/ERK kinase kinase kinase 3) (MEK kinase kinase 3) (MEKKK 3)
DE (Germinal center kinase related protein kinase) (GLK).
GN MAP4K3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), FUNCTION, TISSUE SPECIFICITY, AND
RP MUTAGENESIS OF LYS-48.
RX TISSUE=Macrophage, and Skeletal muscle;
RX MEDLINE=97420743; PubMed=9275185;
RA Diener K., Wang X.S., Chen C., Meyer C.F., Keesler G., Zukowski M.,
RA Tan T.-H., Yao Z.;
RT "Activation of the c-Jun N-terminal kinase pathway by a novel protein
RT kinase related to human germinal center kinase.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9687-9692(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RA Gorry M.C., Zhang Y., Marks J.J., Suppe B., Hart S., Cortelli J.,
RA Pallos D., Hart T.C.;
RT "Physical/genetic map of the 2p22-2p21 region on chromosome 2.";
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-712 FROM N.A. (ISOFORM 1).
RA Edwards J., Wohldmann P., Hawkins M., Harkins R.;
RA Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RL -!- FUNCTION: May play a role in the response to environmental stress.
CC Appears to act upstream of the c-jun N-terminal
CC pathway.
CC -!- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a
CC phosphoprotein.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Interacts with SH3GL2. Interaction appears to regulate
CC MAP4K3-mediated JNK activation (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q81VH8-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=Q81VH8-2; Sequence=VSP_007052;
CC Name=3;
CC IsoId=Q81VH8-3; Sequence=VSP_007053;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in all tissues
CC examined, with high levels in heart, brain, placenta, skeletal
CC muscle, kidney and pancreas and lower levels in lung and
CC liver.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -!- SIMILARITY: Contains 1 CNH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000145; AAC15472.1; -
DR EMBL; AF445413; AAN75849.1; -
DR EMBL; AF445385; AAN75849.1; JOINED.
DR EMBL; AF445386; AAN75849.1; JOINED.
```

DR EMBL; AF445387; AAN75849.1; JOINED.
DR EMBL; AF445388; AAN75849.1; JOINED.
DR EMBL; AF445390; AAN75849.1; JOINED.
DR EMBL; AF445391; AAN75849.1; JOINED.
DR EMBL; AF445392; AAN75849.1; JOINED.
DR EMBL; AF445393; AAN75849.1; JOINED.
DR EMBL; AF445394; AAN75849.1; JOINED.
DR EMBL; AF445395; AAN75849.1; JOINED.
DR EMBL; AF445396; AAN75849.1; JOINED.
DR EMBL; AF445397; AAN75849.1; JOINED.
DR EMBL; AF445398; AAN75849.1; JOINED.
DR EMBL; AF445399; AAN75849.1; JOINED.
DR EMBL; AF445400; AAN75849.1; JOINED.
DR EMBL; AF445401; AAN75849.1; JOINED.
DR EMBL; AF445402; AAN75849.1; JOINED.
DR EMBL; AF445403; AAN75849.1; JOINED.
DR EMBL; AF445404; AAN75849.1; JOINED.
DR EMBL; AF445405; AAN75849.1; JOINED.
DR EMBL; AF445406; AAN75849.1; JOINED.
DR EMBL; AF445407; AAN75849.1; JOINED.
DR EMBL; AF445408; AAN75849.1; JOINED.
DR EMBL; AF445409; AAN75849.1; JOINED.
DR EMBL; AF445410; AAN75849.1; JOINED.
DR EMBL; AF445411; AAN75849.1; JOINED.
DR EMBL; AF445412; AAN75849.1; JOINED.
DR EMBL; AF445413; AAN75850.1; JOINED.
DR EMBL; AF445385; AAN75850.1; JOINED.
DR EMBL; AF445386; AAN75850.1; JOINED.
DR EMBL; AF445387; AAN75850.1; JOINED.
DR EMBL; AF445388; AAN75850.1; JOINED.
DR EMBL; AF445390; AAN75850.1; JOINED.
DR EMBL; AF445391; AAN75850.1; JOINED.
DR EMBL; AF445392; AAN75850.1; JOINED.
DR EMBL; AF445393; AAN75850.1; JOINED.
DR EMBL; AF445394; AAN75850.1; JOINED.
DR EMBL; AF445395; AAN75850.1; JOINED.
DR EMBL; AF445397; AAN75850.1; JOINED.
DR EMBL; AF445398; AAN75850.1; JOINED.
DR EMBL; AF445399; AAN75850.1; JOINED.
DR EMBL; AF445399; AAN75850.1; JOINED.
DR EMBL; AF445400; AAN75850.1; JOINED.
DR EMBL; AF445401; AAN75850.1; JOINED.
DR EMBL; AF445402; AAN75850.1; JOINED.
DR EMBL; AF445403; AAN75850.1; JOINED.
DR EMBL; AF445404; AAN75850.1; JOINED.
DR EMBL; AF445405; AAN75850.1; JOINED.
DR EMBL; AF445406; AAN75850.1; JOINED.
DR EMBL; AF445407; AAN75850.1; JOINED.
DR EMBL; AF445408; AAN75850.1; JOINED.
DR EMBL; AF445409; AAN75850.1; JOINED.
DR EMBL; AF445410; AAN75850.1; JOINED.
DR EMBL; AF445411; AAN75850.1; JOINED.
DR EMBL; AF445412; AAN75850.1; JOINED.
DR EMBL; AC007684; AAF19240.1; JOINED.
DR HSSP; P24941; 1838.
DR Genew; HGNC:6865; MAP4K3.
DR MIM; 604921; -.
DR InterPro; IPR001180; Citron.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR01245; Tyr_pkinase.
DR Pfam; PF00780; CNH; 1.
DR Pfam; PF00069; pkinase; 1.
DR -ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00036; CNH; 1.
DR SMART; SM00220; S-TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
KW ATP-binding; Transferase; Serine/threonine-protein kinase;
KW Alternative splicing.
FT DOMAIN 16 273 PROTEIN_KINASE.
FT DOMAIN 562 874 CNH.
FT NP_BIND 22 30 ATP (BY SIMILARITY).

FT BINDING 48 48 ATP.
FT ACT_SITE 136 136 BY SIMILARITY.
FT VARSPLIC 1 12 MNPFDLSRRNP -> MA (in isoform 2).
FT VARSPLIC 352 372 /FTId-VSP_007052.
FT VARSPLIC 352 372 Missing (in isoform 3).
FT MUTAGEN 48 48 /FTId-VSP_007053.
FT CONFLICT 392 392 K->E: LOSS OF KINASE ACTIVITY AND ABILITY
FT CONFLICT 392 392 TO ACTIVATE JNK FAMILY.
FT SEQUENCE 894 AA; 101315 MW; 6EB77BBB34E5B733 CRC64;
SQ
Query Match 5.88; Score 117.5; DB 1; Length 894;
Best Local Similarity 23.18; Pred. No. 1.1;
Matches 69; Conservative 30; Mismatches 93; Indels 107; Gaps 15;
QY 31 AQLPSSHNPPVSCQMPLLDSNTSHQIMDTPDEEFPNSYLLRACSGPOQASSSGPPNH 90
Db 429 AKIPPLPPPKSIFIP-----QEMHSTED-----NOGTTKRC-----PMSSGSPAK 470
QY 91 HSQSTLRPLP--PPHNHTLSHHSSANSLN-----RNSLTNRSSQIHAPA 134
Db 471 PSQVPPPPPPRLPPHPKPVALNGMSSFLNGERDGSLSLCOQNEHRTNLSRKEKDVPK 530
QY 135 PAPNDLATPE-----SVQLQDSW-----LMSNV 159
Db 531 PISNGLPPTPKVHMGACFSKVFNGCPLKIHCASSWINPDTRDQYLIFGAEEGIYTLNLE 590
QY 160 PLET-----RHLEKTSSTGSGTPLFSSSSPGY-----PLTSGVYTP---195
Db 591 LHETSMELPPRRTWLYVNNCLLSISGRASQLYSHNLPGLFDYARQMKLPVAPAHK 650
QY 196 -PPRLPRTNFTSRKAFKLRK- SKYCSWRKAAL-----SAIAAALLLAILAYFTVP 245
Db 651 LPDRILPRK-FSVSA---KIPETKWCQ-KCCVVRNPVTGHKYLCCGALQTSIVLLEWYEP 704

Search completed: August 26, 2003, 14:31:13
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:03 ; Search time 44 Seconds
(without alignments)
2234.501 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHANKGRPSAEGRP...NDGKDEWSEFNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1889.5	93.7	2764	11 Q9WTS5	Q9WTS5 mus musculus
2	1889.5	93.7	2765	11 Q9RIK2	Q9RIK2 rattus norv
3	1832.5	90.9	831	13 Q9PU49	Q9PU49 gallus gall
4	1820.5	90.3	2802	13 Q9DER5	Q9DER5 gallus gall
5	975.5	48.4	2715	11 Q9WTS6	Q9WTS6 mus musculus
6	917	45.5	2590	13 Q9W7R4	Q9W7R4 brachydanio
7	853	42.3	2825	11 Q70465	Q70465 mus musculus
8	830	41.2	2824	13 Q9W7R3	Q9W7R3 brachydanio
9	793.5	39.3	2771	11 Q9WTS7	Q9WTS7 mus musculus
10	790.5	39.2	560	4 P98202	P98202 homo sapien
11	770	38.2	442	4 Q9NZJ2	Q9NZJ2 homo sapien
12	745.5	37.0	2705	13 Q9W6V6	Q9W6V6 gallus gall
13	739	36.6	536	11 Q8C8D2	Q8C8D2 mus musculus
14	739	36.6	2725	4 Q9UK24	Q9UK24 homo sapien
15	737.5	36.6	2731	11 Q9WTS4	Q9WTS4 mus musculus
16	461.5	22.9	2346	11 Q9JLC1	Q9JLC1 mus musculus

17	427	21.2	159	4	075999	075999 homo sapien
18	207.5	10.3	2731	5	Q9VNU6	Q9VNU6 drosophila
19	207.5	10.3	2731	5	O61307	O61307 drosophila
20	205.5	10.2	2731	5	O18366	O18366 drosophila
21	186.5	9.2	2515	5	O24551	O24551 drosophila
22	177	8.8	1111	10	Q9FH74	Q9FH74 arabidopsis
23	177	8.8	1113	10	O8LPR4	O8LPR4 arabidopsis
24	156.5	7.8	528	6	Q29071	Q29071 sus scrofa
25	154.5	7.7	3004	5	Q24550	Q24550 drosophila
26	154.5	7.7	3004	5	Q9VYN8	Q9VYN8 drosophila
27	144	7.1	309	4	Q9H7C7	Q9H7C7 homo sapien
28	141	7.0	2531	5	Q8MP22	Q8MP22 caenorhabdi
29	141	7.0	2560	5	Q21980	Q21980 caenorhabdi
30	137.5	6.8	325	4	Q8NAK1	Q8NAK1 homo sapien
31	136.5	6.8	1895	5	Q9VWS5	Q9VWS5 drosophila
32	135	6.7	1466	10	Q94HW7	Q94HW7 arabidopsis
33	134.5	6.7	577	4	Q8N811	Q8N811 homo sapien
34	132.5	6.6	2703	5	Q9VEG7	Q9VEG7 drosophila
35	132.5	6.6	2715	5	O61603	O61603 drosophila
36	132.5	6.6	2716	5	Q8IN94	Q8IN94 drosophila
37	132	6.5	421	4	O15329	O15329 homo sapien
38	132	6.5	1268	11	Q63623	Q63623 rattus norv
39	132	6.5	1466	10	Q94HW2	Q94HW2 arabidopsis
40	131	6.5	1892	5	Q97007	Q97007 leishmania
41	129.5	6.4	731	10	O65530	O65530 arabidopsis
42	129	6.4	383	5	Q9Y075	Q9Y075 leishmania
43	129	6.4	1330	10	Q9SXQ2	Q9SXQ2 arabidopsis
44	129	6.4	1456	10	O92T94	O92T94 arabidopsis
45	128.5	6.4	518	3	Q8NJ53	Q8NJ53 candida alb

ALIGNMENTS

RESULT 1

Q9WTS5 Q9WTS5 PRELIMINARY; PRT; 2764 AA.
AC Q9WTS5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Ten-m2.
GN OD22 OR TEN-M2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=Brain;
RA Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,
RA Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;
RT "Mouse Ten-m/Odz is a new family of dimeric type II transmembrane
RT proteins expressed in many tissues."
RL J. Cell Biol. 0:0-0(1999).
DR EMBL; AB025411; BAA77397.1; -
DR HSP; P35555; 1ENN.
DR MGD; MGI:1345184; Odz2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001258; NHL.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF01436; NHL; 6.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00181; EGF; 5.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain.
SQ SEQUENCE 2764 AA; 306464 MW; 73BA3D916D0F0344 CRC64;

Query Match 93.7%; Score 1889.5; DB 11; Length 2764;
 Best Local Similarity 88.5%; Pred. No. 6.3e-161;
 Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 69
 DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQOASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSLNRRSQ 129
 DB 222 SYLLRACSGPQOASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSLNRRSQ 281

QY 130 IHAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 189
 DB 282 IHAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 341

QY 190 GVIYTPPPRLPNTFSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFI----- 243
 DB 342 GVIYTPPPRLPNTFSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFIAMHLLG 401

QY 244 -----VPWSLKNSSIDSGAEVGRRTQEV 268
 DB 402 LNWQLPADGHTFNNGVTRGLPGNDVAVTPSGKVPWSLKNSSIDSGAEVGRRTQEV 461

QY 269 PGVFWRSQIHISQOFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 328
 DB 462 PGVFWRSQIHISQOFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 521

QY 329 ESPRERSIQTIVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 378
 DB 522 ESPRERSIQTIVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 571

RESULT 2

Q9R1K2 ID Q9R1K2 PRELIMINARY; PRT; 2765 AA.
 AC Q9R1K2
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Neurestin alpha.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP "Neurestin: putative transmembrane molecule implicated in neuronal development."
 RL Dev. Biol. 212:165-181(1999).
 DR EMBL; AF086607; AAD47383.1; .
 DR HSSP; P00750; ITPG.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 5.
 DR TIGRfams; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2765 AA; 306568 MW; 6748D70D5FFD8F0E CRC64;

Query Match 93.7%; Score 1889.5; DB 11; Length 2765;
 Best Local Similarity 88.5%; Pred. No. 6.3e-161;
 Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 69

DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQOASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSLNRRSQ 129
 DB 222 SYLLRACSGPQOASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSLNRRSQ 281

QY 130 IHAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 189
 DB 282 IHAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 341

QY 190 GVIYTPPPRLPNTFSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFI----- 243
 DB 342 GVIYTPPPRLPNTFSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFIAMHLLG 401

QY 244 -----VPWSLKNSSIDSGAEVGRRTQEV 268
 DB 402 LNWQLPADGHTFNNGVTRGLPGNDVAVTPSGKVPWSLKNSSIDSGAEVGRRTQEV 461

QY 269 PGVFWRSQIHISQOFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 328
 DB 462 PGVFWRSQIHISQOFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 521

QY 329 ESPRERSIQTIVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 378
 DB 522 ESPRERSIQTIVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 571

RESULT 3

Q9P049 ID Q9P049 PRELIMINARY; PRT; 831 AA.
 AC Q9P049
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Teneurin-2 protein.
 GN TEN2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20057765; PubMed-10588872;
 RA Rubin B.P., Tucker R.P., Martin D., Chiquet-Ehrismann R.;
 RT "Teneurins: a novel family of neuronal cell surface proteins in
 RT vertebrates, homologous to the Drosophila pair-rule gene product Ten-
 RT m.";
 RL Dev. Biol. 216:195-209(1999).
 DR EMBL; AJ245711; CAB57257.1; .
 DR HSSP; P00750; ITPG.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 4.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 5.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 6.
 KW EGF-like domain.
 SQ SEQUENCE 831 AA; 90359 MW; F166547445EAF7D2 CRC64;

Query Match 90.9%; Score 1832.5; DB 13; Length 831;
 Best Local Similarity 85.1%; Pred. No. 1.8e-156;
 Matches 349; Conservative 9; Mismatches 11; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 69
 DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQOASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSLNRRSQ 129

Db 256 TCGTTPLESTATPGYTMASGVSPPTPLPNTLSRSFAFKKSKYCSWRCALCAV 315
 QY 230 AAALLAILLAYFI-----VPWSL----- 248
 Db 316 GVSVLLAILLSYFIAMHFLGNLHQOTENDTFENGKVNSDTPTNTVSLPSGDNGKLGG 375
 QY 249 ---KNSIDSGEAEVGRRTQEVPPGFWRSQTHISQPLKFNISLQKDALRGVYIRRG 305
 Db 376 FTHENTIDSGEDIGIRRAIQEYPPGFIWRSOLFIDQPOPLKFNISLQKDALIGVYGRKG 435
 QY 306 LPSPHAYDPMERLDG-----KFKWSVSPRRRSSTOTLVQNEAVFQYLDVGLWHLAF 360
 Db 436 LPSPHAYDPMERLDG-----KFKWSVSPRRRSSTOTLVQNEAVFQYLDVGLWHLAF 360
 QY 361 YNDGKDKEMVSFNTVYLDGTI 381
 Db 496 YNDGKNPEQVSFNTIVIESVV 516

RESULT 6

Q9W7R4 PRELIMINARY; PRT; 2590 AA.
 AC Q9W7R4;
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Ten-m3.
 GN ODZ3 OR TENM3 OR TEN-M3.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99425191; PubMed=10495292;
 RA Mieda M., Kikuchi Y., Hirate Y., Aoki M., Okamoto H.;
 RT "Compartmentalized expression of zebrafish ten-m3 and ten-m4,
 RT homologues of the Drosophila tenm /odd Oz gene, in the central nervous
 RT system.";
 RL Mech. Dev. 87:223-227(1999).
 DR EMBL; AB026979; BAA81892.1; -.
 DR HSP; P10968; 7WGA.
 DR ZFIN; ZDB-GENE-990714-19; odz3.
 DR InterPro; IPR006210; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 6.
 DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2590 AA; 288586 MW; 597592866210148D CRC64;

Query Match 45.5%; Score 917; DB 13; Length 2590;
 Best Local Similarity 46.5%; Pred. No. 3.2e-73;
 Matches 190; Conservative 61; Mismatches 94; Indels 64; Gaps 9;
 QY 23 SPSPILPSAQLPSSNPPPV-----SCQPLLDNNTSHQIMDNDDEEFPNSVLLRACSG 78
 Db 6 SPSPV-----TEHSHQPPSPNLHNDQSSILSNATTAQVQSDSEETAVLYRPVYQPA 60
 QY 79 PQQASSGPPNHH--SQSLRPLPPPHNHTLSHHSSANSLSNLSITNRRSQIHA-PAP 135
 Db 61 PSHSCNEQSNHQQCSQSLT-PPVPPPHK-----QPSTVALNHNLSRRNVSPAPPA 114
 QY 136 APNDLATTPEVQLQDSWLVNSVPLETRHFLKTSSTGTPLFSSSSSGYPLTSGVTYTP 195
 Db 115 LPALQTTPESVPLQDSWLVNSVPLESRHFLKTSSTGTPLEFATPGYMATGAYSP 174

QY 196 PPRLPNTFSRAFLKPKSKYCSWKAALSIAAALLAILLAY-----FIVPWSL- 248
 Db 175 PTRPLPNTLSRAFAFKKSKYCSWRCALTASAMAVSILLSVLLCYCIAMHFLGNLQ 234
 QY 249 -----KNSIDSGEAEVGRRTQEVPPGFW 274
 Db 235 ETEGYAFENGQVKSSTATNAVLTALSTENKVPYQENNTIDTGEVDVGRRAVDVPPGTFW 294
 QY 275 RSQIHISQOPFLKFNISLQKDALRGVYIRRGFLPSPHAYDPMERLDG-----KFKWSV 329
 Db 295 RTOLFIDQPSLKFNSVQRGALVGYGRKGLPPTHQYDFVELLDGSLIAEKKRGLVE 354
 QY 330 SPERSIQTLVQNEAVFQYLDVGLWHLAFYNDGDKEMVSFNTVYLD 378
 Db 355 VEGAARKARSVNVHEAEFIRFLDSGTWHLAFYNDGKNAEQSYNTIID 403

RESULT 7

O70465 PRELIMINARY; PRT; 2825 AA.
 AC O70465;
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE DOC4.
 GN NRGI OR ODZ4 OR DOC4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIH/Swiss;
 RX MEDLINE=98315054; PubMed=9649432;
 RA Wang X.Z., Kuroda M., Sok J., Batchvarova N., Kimmel R., Chung P.,
 RA Zinszner H., Ron D.;
 RT "Identification of novel stress-induced genes downstream of chop.";
 RL EMBO J. 17:3619-3630(1998).
 DR EMBL; AF059485; AAC31807.1; -.
 DR HSP; P16109; IFSB.
 DR MGD; MGI:96083; Nrql.
 DR MGD; MGI:2447063; Odz4.
 DR InterPro; IPR006210; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 6.
 DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2825 AA; 313483 MW; A120D98080886032 CRC64;

Query Match 42.3%; Score 853; DB 11; Length 2825;
 Best Local Similarity 40.1%; Pred. No. 2.1e-67;
 Matches 193; Conservative 55; Mismatches 111; Indels 122; Gaps 12;
 QY 12 SAEGAPRIPPTSPSLLPSAQLPSSNPPPV-----NHSQSLRPLPPPHNHTLSHH--SSANSLSNRLSLN 125
 Db 161 NTEGTAPL-HCSSASSTPIEQSPSPPPSPANESQRRLLGNGVAQTPDSDSEEEFVPS 219
 QY 71 YLIRACSGPQQAASGPP-----NHSQSLRPLPPPHNHTLSHH--SSANSLSNRLSLN 125
 Db 220 FLVKSSASLGVANDHPSSLQNHPLRT--PPPLPHANTPNQHAASLSNLRGNFTP 277
 QY 126 RRSQIHAPAPND-----LATTPEVQLQDSWLVNSVPLETR----- 164
 Db 278 RSN-----PSPAPTDHLSGEPAPGSAQEPHQAQDNVLSKIPVETRNLGKQPLFGTWD 333
 QY 165 -----HFLFKYSSGSTPLFSSSSPGYPLTSGVTYTPPRLPRLPNT 204


```

QY 69 NSYLLRACSGPQOAGSSGPPNHHQSOTLRPLPP-PHNHTLSHHH-SSANSNRNSLTNR 126
DB 149 NSNLTLTDTHEHENTETDHPSSSLQNHPRLRTPPPPLPHAHTPNQHHASINSLNRGNFTPR 208
QY 127 RSQIHAPAPAPND-----LATTESVOLQDSWVNSNVPLETR----- 164
DB 209 SN-----PSPAPTDHSLSGEPAGAGAEPTTHAQDNWLLNSNIPLERNLGKQPFGLTLDQN 264
QY 165 -----HFLKTSSTGSPFLSSSSPGYPLTSGTYTTPPPRLPRNTF 205
DB 265 LIEMDILSASRDGAYSDGHFLFK-PGTSPLCTTSPGYPLTSSVYSPPPRLPRSTF 323
QY 206 SRKAFLLKPKSKYCKWKAALSIAAALLAILLAYFI-----VPSWL- 248
DB 324 SRPAFLNLKPKSKYCKWKAALSIAISATVILLAYFVAMHLFGLNHLQPMQOMYE 383
QY 249 -----KNSSIDSGEAEVGR 263
DB 384 ITEDTASSWPVPTDVSILPSGGTGLTDPDRKGAGAEKSPSSFLPEDSFIDSGEIDVGR 443
QY 264 VTQEPVPGVWRSQIHISQPKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK- 322
DB 444 ASQKIPPGTFRWSQVFDHPVHLKFNYSGLKAAALVGIYGRKGLPPSHQTFDFVELLDGR 503
QY 323 -----EKWSVSPSPRRRSIOTLVQNEAVFVQYLDVGLWHLAFYNDGKEMVSFNTVLD 378
DB 504 LTTQARSLEGPPQORSGVPPSPSHSETGTGFIQYLDGSIWHLAFYNDGKESEVVSFLATAIE 563

RESULT 10
P98202
ID P98202 PRELIMINARY; PRT; 560 AA.
AC P98202; 014667;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Downstream of CHOP protein (DOC-4) (Teneurin) (Fragment).
GN OD24 OR DOC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. OF GAMMA-HERGULIN FUSION PROTEIN.
RC TISSUE-BREAST CANCER;
RX MEDLINE-97472144; PubMed-9333014;
RA Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;
RT "Gamma-hergulin: a novel heregulin isoform that is an autocrine
RT growth factor for the human breast cancer cell line, MDA-MB-175.";
RL Oncogene 15:1385-1394 (1997).
RN [2]
RP CHROMOSOMAL TRANSLOCATION.
RX MEDLINE-99455251; PubMed-10523851;
RA Wang X.-Z., Jolicoeur E.M., Conte N., Chaffanet M., Zhang Y.,
RA Mozziconacci M.-J., Feiner H., Birnbaum D., Pebusque M.-J., Ron D.;
RT "Gamma-hergulin is the product of a chromosomal translocation fusing
RT the DOC4 and HGL/NGR1 genes in the MDA-MB-175 breast cancer cell
RT line.";
RL Oncogene 18:5718-5721 (1999).
RN [3]
RP CHROMOSOMAL TRANSLOCATION.
RX MEDLINE-20065180; PubMed-10597312;
RA Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;
RT "Gamma-hergulin: a fusion gene of DOC-4 and neuregulin-1 derived from
RT a chromosome translocation.";
RL Oncogene 18:7110-7114 (1999).
RN [4]
RP SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC 1- INDUCTION: BY STRESS.
CC 1- DISEASE: INVOLVED IN A RARE T(8;11) CHROMOSOMAL TRANSLOCATION THAT
CC FUSES THE 5' END OF OD24 TO NRG1 (ISOFORM BETA3). THE PRODUCT OF
CC THIS TRANSLOCATION HAS FIRST BE THOUGHT TO BE AN ALTERNATIVE
CC SPLICED ISOFORM, CALLED GAMMA-HERGULIN. NOT DETECTED IN BREAST
CC CARCINOMA SAMPLES, INCLUDING DUCTAL, LOBULAR, MEDULLARY, AND

```

```

CC MUCINOUS HISTOLOGICAL TYPES, NEITHER IN OTHER BREAST CANCER CELL
CC LINES.
CC 1- SIMILARITY: TO D. MELANOGASTER TENN/ODZ.
DR EMBL: AF009227; AAC51756.1; ALT TERM.
KW Transmembrane; Chromosomal translocation.
FT DOMAIN 179 182 POLY-PRO.
FT TRANSMEM 346 366 POTENTIAL.
FT NON_TER 560 560
SQ SEQUENCE 560 AA; 61494 MW; 3927565E76C26C1D CRC64;

Query Match 39.2%; Score 790.5; DB 4; Length 560;
Best Local Similarity 41.5%; Pred. No. 1.1e-62;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

QY 69 NSYLLRACSGPQOAGSSGPPNHHQSOTLRPLPP-PHNHTLSHHH-SSANSNRNSLTNR 126
DB 149 NSNLTLTDTHEHENTETDHPGQLNHARLRTPPPPLSHAHTPNQHHASINSLNRGNFTPR 208
QY 127 RSQIHAPAPAPND--LATTP-----ESVOLQDSWVNSNVPLETR----- 164
DB 209 SN-----PSPAPTDHSLSGEPAGAGAEPTTHAQDNWLLNSNIPLERNLGKQPFGLTLDQN 264
QY 165 -----HFLKTSSTGSPFLSSSSPGYPLTSGTYTTPPPRLPRNTF 205
DB 265 LIEMDILSASRDGAYSDGHFLFK-PGTSPLCTTSPGYPLTSSVYSPPPRLPRSTF 323
QY 206 SRKAFLLKPKSKYCKWKAALSIAAALLAILLAYFI-----VPSWL- 248
DB 324 SRPAFLNLKPKSKYCKWKAALSIAISATVILLAYFVAMHLFGLNHLQPMQOMYEIT 383
QY 249 -----KNSSIDSGEAEVGR 265
DB 384 EDTASSWPVPTDVSILPSGGTGLTDPDRKGKTEGKSPSSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVWRSQIHISQPKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK--- 322
DB 444 QKIPPGTFRWSQVFDHPVHLKFNYSGLKAAALVGIYGRKGLPPSHQTFDFVELLDGRLL 503
QY 323 --EKWSVSPSPRRRSIOTLVQNEAVFVQYLDVGLWHLAFYNDGKEMVSFNTVVL 377
DB 504 TQARSLEGTPROSGRTVPPSPSHSETGTGFIQYLDGSIWHLAFYNDGKESEVVSFLTAI 560

RESULT 11
Q9NZJ2
ID Q9NZJ2 PRELIMINARY; PRT; 442 AA.
AC Q9NZJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OD23 (Fragment).
GN OD23.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP Ben-Zur T., Feige E., Motro B., Wides R.;
RT "The mammalian Odz gene family: Homologs of a Drosophila pair rule
RT gene with expression implying distinct yet overlapping developmental
RT roles.";
RL Dev. Biol. 117:107-120 (2000).
DR EMBL: AF195420; AAF28318.1; -.
FT NON_TER 442 442
SQ SEQUENCE 442 AA; 49073 MW; F4DD0DA48E66B75 CRC64;

Query Match 38.2%; Score 770; DB 4; Length 442;
Best Local Similarity 50.8%; Pred. No. 5.5e-61;
Matches 156; Conservative 36; Mismatches 55; Indels 60; Gaps 5;

QY 51 SNTSHQIMDTNPDEEESPNSYLLRACSGPQOAGSSGPPNHHQSOTLRPLPPPHNHTLSH 110

```

```

Db 151 SNSALTLTDEHENK-----SDSENEQASNOGOSTLQ-PLPPSHKQHSQAQ 195
QY 111 HSSANSLNRSNLTNRSSQIHA-PAPAPNDLATTTPESVOLQDSWVLSNVPLETRHFLFK 169
Db 196 HPSITSLNRSNLTNRNQSPPAAPALPAELQTTTPESVOLQDSWVLSNVPLESRHFLFK 255
QY 170 TSSGSPLESSSPGVLTSVYVTPPPRLLPRNTFSRKAFLKPKSKYCSWKCAALSAL 229
Db 256 TGTGTPPSTATPGYTMASSGVSYPTRPLPRNTLSRSFAFKKSKYCSWKCTALCAV 315
QY 230 AAALLAILLAYFI-----VPSWL----- 248
Db 316 GVSVLAAILLSFIAHHLGLNQLQOQTENDTFENGKVSNDTMTVTVSLPSGDNKGLGG 375
QY 249 ----KNSIDSGAEVGRVYQVPPGVFWRSQIHISQPFKFNISLGDALFGVYIRRG 305
Db 376 FTQENTIDSGELDIGRRAIQEIPPGIFWRSOLFIDQPFKFNISLQKDALIGVYGRKK 435
QY 306 LPPSHQAQ 312
Db 436 LPPSHQAQ 442

RESULT 12
Q9W6V6 PRELIMINARY; PRT; 2705 AA.
AC Q9W6V6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Teneurin-1.
GN TEN-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99276585; PubMed=10341219;
RA Minet A.D., Rubin B.P., Tucker R.P., Baumgartner S.,
RA Chiquet-Ehrismann R.;
RT "Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene
RT ten-m, is a neuronal protein with a novel type of heparin-binding
RT domain.";
RL J. Cell Sci. 112:2019-2032(1999).
DR EMBL; AJ238613; CAB43098.1; -
DR HSSP; P00750; ITPG.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 5.
DR SMART; SM00181; EGF; 6.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain.
SQ SEQUENCE 2705 AA; 302385 MW; 230F03D1999037D2 CRC64;

Query Match 37.0%; Score 745.5; DB 13; Length 2705;
Best Local Similarity 39.1%; Pred. No. 1e-57;
Matches 174; Conservative 56; Mismatches 120; Indels 95; Gaps 15;

QY 2 LHAANKRPSAEAGRIPTSPSLPSAQL-----PSSNPPVSCOMPLLDNSVS 54
Db 91 LHGVSEHSYP-LEVGSVDVTETEGGASPDHALRMWRGMKSEHS-----LSSRAN 141
QY 55 HOIMDNPDEEFPNSVLLRACSGPQQAASSGPPNHHOSOTLR---PPLPPPHNHTLSHH 111
Db 142 SALSLSLTDHE-----RKSDG-----ENDMPGSPHNQFTFRPLPPPPPHACTCTRK 189

```

```

QY 112 -HSSANSLNRSNLTNRSSQIHA-PAPAPNDLATTTPESVOLQDSWVLSNVPLETRHFLFK 169
Db 190 PPAADSLQRRSMTT-RSQ---PSPAAPTPTQDSVHLHNSWVLSNIPLETRHFLFK 245
QY 170 TSSGSPLESSSPGVLTSVYVTPPPRLLPRNTFSRKAFLKPKSKYCSWKCAALSAL 229
Db 246 HGSSSAIFSAASQNYPLTSNTVYSPRPLPRNTFSRPAFTFSKPYRCNNKCTALSAT 305
QY 230 AAALLAILLAYFI-----VPSWL----- 248
Db 306 ATVTVLLALLAYVIAVHLFGLTWQLQPVGOLYENGSKNGKAESTDDTTSYGKGVSD 365
QY 249 ----KNSIDSGAEVGRVYQVPPGVFWRSQIHISQPFKFNISLGDALFGV 300
Db 366 KTEKKVFQGRAIDTGEVEIGAQMOTIPPGFLFWRFQITHHVPLVKFNISLAKDSILGI 425
QY 301 YTRGLPPSHAQYDFMERLDGKREKSVSPRRRSIQTLVN-----EAVFYQYLDV 353
Db 426 YGRRNIPPTHTQDFVKLMDGKQ--LIKQEPKNSEEPQAAPRNLILTSLOETGFIEYMDQ 483
QY 354 GLWHLAFYNDGKDKEMVSENTVVLD 378
Db 484 GAHHMAFYNDGKKVEQVFLTTAIE 508

RESULT 13
Q8C8D2 PRELIMINARY; PRT; 536 AA.
AC Q8C8D2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Odd Oz/ten-m homolog 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK047423; BAC33054.1; -
SQ SEQUENCE 536 AA; 60081 MW; C9814627AE9B363F CRC64;

Query Match 36.6%; Score 739; DB 11; Length 536;
Best Local Similarity 41.4%; Pred. No. 4.5e-58;
Matches 170; Conservative 50; Mismatches 111; Indels 80; Gaps 11;

QY 33 LPSSHNPPVSCOMPLLDNSNTSHQIMDTPDEE-----FSPNSVLLRACSGPQA 82
Db 128 MKSEHS-----SC-----LSSRANSALSLTDHDKRGKGFSPVCCDMEAPDASQD 179
QY 83 SSSGPPNHHQSOTLRPLPPPHNHTLSHH-HSSANSLNRSNLTNRSSQIHA-PAPNDLA 141
Db 180 MOSSPHNQFTFRPLPPPPPHACTCARPPPTVDSLQRRSMTT-RSQSPAPAPP--T 236
QY 142 TPESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGVTYPPRLLP 201
Db 237 STQDSVHLHNSWVLSNIPLETRHFLFKHSGSSAIFSAASQNYPLTSNTVYSPRPLP 296
QY 202 RNTFSRKAFLKPKSKYCSWKCAALSALAILLAYFYV-----PMSL----- 248
Db 297 RSTFSRPAFTFNKPYRCNNKCTALSATAITVTALLAYVIVHLFGLTWQLQPVQGIYA 356
QY 249 -----KNSIDSGAEVGRVYQVPPGVFWRSQIHISQPFKFNISLGDALFGV 274
Db 357 NGISNGNPGCTESMDTYSPIGGRVSDKSEKVKFGKRAIDTGEVDIGAQVMTIPPLEW 416

```

QY	275	RSQIHISQPOFLKFNISLGDALFVGYVIRGLPPSHAGYDFMRDLGRK-----KWS 326
DB	417	RFQITTHPIYLFKNISLAKDSLGIYGRNIPPTHTQDFVFKLMDGKQLVKQDKSSDD 476
QY	327	WVSPERRRSIQTLQVNEAVFQVYLDVGLGHLAFYNDGDKREMFSEFNVVL 377
DB	477	IQHSP--RNILTSLQ-ETGFIEYDQGPWYLAFLAFNDGKKEQVFLTAI 524
RESULT 14		
Q9UK24		
ID	Q9UK24	PRELIMINARY; PRT: 2725 AA.
AC	Q9UK24	
DT	01-MAY-2000	(Tremblrel. 13, Created)
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)
DT	01-MAR-2003	(Tremblrel. 23, Last annotation update)
GN	Tenascin-M1.	
GN	TNN1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20025749; PubMed=10536288;	
RA	Brandau O., Schuster V., Weiss M., Hellebrand H., Fink F.M.,	
RA	Kreczy A., Friedrich W., Strahm B., Niemeyer C., Belohradsky B.H.,	
RA	Meindl A.;	
RT	*Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are	
RT	mutated in the SH2B1A gene, as are patients with X-linked	
RT	lymphoproliferative disease (XLP).";	
RL	Hum. Mol. Genet. 8:2407-2413(1999).	
RL	EMBL; AF100772; AAF04723.1; .	
DR	HSSP; P00750; ITPG.	
DR	Genew; HGNC:8117; ODZ1.	
DR	InterPro; IPR006209; EGF_like.	
DR	InterPro; IPR006210; IEGF.	
DR	InterPro; IPR001258; NHL.	
DR	InterPro; IPR000209; Peptidase_S8.	
DR	InterPro; IPR006530; YD.	
DR	Pfam; PF00008; EGF; 5.	
DR	Pfam; PF01436; NHL; 6.	
DR	SMART; SM00181; EGF; 6.	
DR	TIGRfams; TIGR01643; YD_repeat_2x; 7.	
DR	PROSITE; PS00022; EGF_1; 8.	
DR	PROSITE; PS01186; EGF_2; 7.	
DR	PROSITE; PS00136; SUBTILASE_ASP; 1.	
KW	EGF-like domain.	
QY	SEQUENCE 2725 AA; 304892 MW; 6F383E54C1B2E1B1 CRC64;	
Query Match		
Best Local Similarity 36.6%; Score 739; DB 4; Length 2725;		
Matches 170; Conservative 51; Mismatches 113; Indels 78; Gaps 10		
QY	33	LPSSHNPVYSCOMPLDSDTSHQIMDNTDEE-----FSPSYLLRACSGPOQA 82
DB	128	MKSEHS-----SC-----LSSRANSALSLTDTHRKSDGNGFKFSPVCCDMEAQAGSQD 179
QY	83	SSGPPNHHSQSTLRPLPPPHNITLSHH--HSSANSLNRLNRRSIOIHAPAPNDLA 141
DB	180	VQSSPNHQFTFRLPPPPPPHACTCAKPPPAADSLQRSSMTT--RSQSPSAAPAPP--T 236
QY	142	TPFESVOLQDSWLNSNVLPETHFLPKTSQSGTFLPSSSPGYPLSGTYTTPPRLLP 201
DB	237	STQDSVHLHNSWLNSNIPLETRHSLFKHSGSSAIFSAASQNYPLTSNTVYSPPPRLP 296
QY	202	RNTFSRAKFLKPKSKYCSWKCAALSIAAALLAILLAYFI-----VPSNL----- 248
DB	297	RSTFSRPAFTFNKPYRCNKKCFALSATAITVTIALLLAYVIAVHLFCLTQLOPVEGEL 356
QY	249	-----KNSSIDSGAEVGRRTQVEVPGV 272

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:58 ; Search time 21 Seconds
(without alignments)
767.640 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017
Sequence: 1 MLHANKRPSAAGRP...NDGDKEMVSFNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	790.5	39.2	560	3	US-08-891-845-4
2	790.5	39.2	560	4	US-09-514-573-4
3	790.5	39.2	768	3	US-08-891-845-2
4	790.5	39.2	768	4	US-09-514-573-2
5	658	32.6	501	3	US-08-891-845-10
6	658	32.6	501	4	US-09-514-573-10
7	122	6.0	872	1	US-08-491-357-3
8	122	6.0	872	3	US-08-968-633-3
9	122	6.0	872	3	US-09-190-466-3
10	122	6.0	872	5	PCT-US96-10823-3
11	117	5.8	612	4	US-08-894-454-163
12	117	5.8	657	1	US-08-264-534-34
13	117	5.8	657	1	US-08-083-590A-13
14	117	5.8	657	1	US-08-465-500-34
15	117	5.8	657	2	US-08-346-128-34
16	117	5.8	657	3	US-08-532-384-13
17	117	5.8	657	3	US-08-893-828-34
18	117	5.8	2471	1	US-08-185-432-16
19	117	5.8	2471	1	US-08-083-590A-19
20	117	5.8	2471	3	US-08-532-384-19
21	117	5.8	2471	4	US-08-899-232-1
22	116.5	5.8	681	1	US-08-083-590A-18
23	116.5	5.8	681	2	US-08-346-128-37
24	116.5	5.8	681	3	US-08-532-384-18
25	116.5	5.8	1078	1	US-08-264-534-32
26	116.5	5.8	1078	1	US-08-083-590A-11
27	116.5	5.8	1078	1	US-08-465-500-32

28	116.5	5.8	1078	2	US-08-346-128-32	Sequence 32, Appl
29	116.5	5.8	1078	3	US-08-532-384-11	Sequence 11, Appl
30	116.5	5.8	1078	3	US-08-893-828-32	Sequence 32, Appl
31	116.5	5.8	2556	1	US-08-083-590A-20	Sequence 20, Appl
32	116.5	5.8	2556	3	US-08-532-384-20	Sequence 20, Appl
33	116	5.8	2972	4	US-09-579-181-2	Sequence 2, Appl
34	116	5.8	3118	4	US-09-579-181-1	Sequence 1, Appl
35	113.5	5.6	1481	2	US-08-616-844-40	Sequence 40, Appl
36	113.5	5.6	1481	2	US-08-599-654-40	Sequence 40, Appl
37	113.5	5.6	1481	3	US-08-944-868A-40	Sequence 40, Appl
38	113.5	5.6	1481	3	US-08-944-423A-40	Sequence 40, Appl
39	113.5	5.6	1481	3	US-08-944-486-40	Sequence 40, Appl
40	113	5.6	1706	2	US-08-459-568-2	Sequence 2, Appl
41	113	5.6	1706	2	US-08-399-411-2	Sequence 2, Appl
42	113	5.6	1706	3	US-08-516-859A-2	Sequence 2, Appl
43	113	5.6	1706	4	US-09-586-472-2	Sequence 2, Appl
44	113	5.6	1706	4	US-09-528-706-2	Sequence 2, Appl
45	112	5.6	472	3	US-08-348-518C-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-891-845-4
; Sequence 4, Application US/08891845
; Patent No. 6096873
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Herregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-891-845-4

Query Match 39.2%; Score 790.5; DB 3; Length 560;
Best Local Similarity 41.5%; Pred. No. 5e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
Qy 69 NSYLLRACSGQQSSGPPNHHQSOTLRPLPP-PHNHTLSHHH;SSANSLNRLTNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQNHARLTPPPPLSHARTPNQHHAAINSLNRLGNFTPR 208

Query Match 39.2%; Score 790.5; DB 3; Length 768;
Best Local Similarity 41.5%; Pred. NO. 7.8e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

US-09-514-573-2

QY 69 NSYLRLACSGPQQAASSGPPNHHQSOTLRPPLPP-PHNHTLSHHH-SSANSLNRLNRLNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQHARLRTPPPLSHAHTPNQHHAAASINSLNRGNFTPR 208
QY 127 RSQIHAPAPAND--LATTP-----ESVQLQDSWVLSNSVPLETR----- 164
DB 209 SN-----PSPAPTHSLSGEPAGGAQEPAAHAQENWLLNSNIPLETRNLGKQPFGLTLODN 264
QY 165 -----HFLFTSSGTPFLFSSSPGYPLTSGVTYTPPPRLLPRTNF 205
DB 265 LIEMDILGASRDGAYS DGHFLEK-PGCTSPFLCTTSPGYPLTSTVYSPPPRLPRSTF 323
QY 206 SRKAFKLKPKSKYCWKCAALSATAAALLAILLAYFI-----VPSWL----- 248
DB 324 ARPAPLKKPKSKYCNWKAALSATVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383
QY 249 -----KNSSIDSGEAEVGRRTV 265
DB 384 EDTASSWPVPTDVSYPGSGTGLETPDRKKGTTGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSQIHISQPOFLKFNISLGDALFGVYIRGGLPPSHAQYDFMERLDGK--- 322
DB 444 QKIPPGTFWRSQVFDHVPVHLKFNVSGLKAALVGIYGRKGLPPSHTQPDFVELLDGRLL 503
QY 323 --EKWSVSPRERSIQTLYQNEAVFVQYLDVGLWHLAFYNDGKDKEMVSFNTVWL 377
DB 504 TQEARSLGTPRQSRGTVPPSSSHETGFIQYLDSDGIWHLAFYNDGKESEVVSFLTTAI 560

RESULT 4

US-09-514-573-2
; Sequence 2, Application US/09514573
; Patent No. 6500941
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Heregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,573
; FILING DATE: 28 FEB 2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/891845
; FILING DATE: 10 JULY 1997
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 768 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-514-573-2

Query Match 39.2%; Score 790.5; DB 4; Length 768;
Best Local Similarity 41.5%; Pred. NO. 7.8e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

QY 69 NSYLRLACSGPQQAASSGPPNHHQSOTLRPPLPP-PHNHTLSHHH-SSANSLNRLNRLNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQHARLRTPPPLSHAHTPNQHHAAASINSLNRGNFTPR 208
QY 127 RSQIHAPAPAND--LATTP-----ESVQLQDSWVLSNSVPLETR----- 164
DB 209 SN-----PSPAPTHSLSGEPAGGAQEPAAHAQENWLLNSNIPLETRNLGKQPFGLTLODN 264
QY 165 -----HFLFTSSGTPFLFSSSPGYPLTSGVTYTPPPRLLPRTNF 205
DB 265 LIEMDILGASRDGAYS DGHFLEK-PGCTSPFLCTTSPGYPLTSTVYSPPPRLPRSTF 323
QY 206 SRKAFKLKPKSKYCWKCAALSATAAALLAILLAYFI-----VPSWL----- 248
DB 324 ARPAPLKKPKSKYCNWKAALSATVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383
QY 249 -----KNSSIDSGEAEVGRRTV 265
DB 384 EDTASSWPVPTDVSYPGSGTGLETPDRKKGTTGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSQIHISQPOFLKFNISLGDALFGVYIRGGLPPSHAQYDFMERLDGK--- 322
DB 444 QKIPPGTFWRSQVFDHVPVHLKFNVSGLKAALVGIYGRKGLPPSHTQPDFVELLDGRLL 503
QY 323 --EKWSVSPRERSIQTLYQNEAVFVQYLDVGLWHLAFYNDGKDKEMVSFNTVWL 377
DB 504 TQEARSLGTPRQSRGTVPPSSSHETGFIQYLDSDGIWHLAFYNDGKESEVVSFLTTAI 560

RESULT 5

US-08-891-845-10
; Sequence 10, Application US/08891845
; Patent No. 6096873
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Heregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994

```
;
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-08-891-845-10
;
; Query Match 32.6%; Score 658; DB 3; Length 501;
; Best Local Similarity 47.5%; Pred. No. 5.8e-52;
; Matches 132; Conservative 33; Mismatches 47; Indels 66; Gaps 4;
;
; QY 165 HFLFKTSSGSTPLFSSSSPGYPLTSGTYVTPPRLPRNTFSKAFKLLKPKSKYCSWKCA 224
; DB 17 HELFK-PGTSPLFCTSPGYPLTSGTYVTPPRLPRNTFSKAFKLLKPKSKYCSWKCA 75
;
; QY 225 ALSATAAALLLAILLAYFI-----VPWSL----- 248
; DB 76 ALSAIVISATLVILLAYFVAMHLFGLNHLQPMEGOMYEITDASSWPVPTDVSILPSG 135
;
; QY 249 -----KNSSIDSGAEVGRVTVQVPPGVFWRSQIHISQPO 284
; DB 136 GTGLETPDRKGKGTGKPPSSFPEDSFIDSGEIDVGRRASQKIPPGTFWRQVFDHPV 195
;
; QY 285 FLKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK-----EKWSVYESPRRSIQT 339
; DB 196 HLKFNYSLGAALVGIYGRKGLPPSHQAQYDFMERLDGK-----EKWSVYESPRRSIQT 339
;
; QY 340 LVQNEAVFYQYLDVGLWHLAFYNDGKEMVSEFNTVVL 377
; DB 256 PSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAT 293
```

RESULT 6

```
US-09-514-573-10
; Sequence 10, Application US/09514573
; Patent No. 6500941
;
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Herregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,573
; FILING DATE: 28 FEB 2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/891845
; FILING DATE: 10 JULY 1997
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-09-514-573-10
;
; Query Match 32.6%; Score 658; DB 4; Length 501;
; Best Local Similarity 47.5%; Pred. No. 5.8e-52;
; Matches 132; Conservative 33; Mismatches 47; Indels 66; Gaps 4;
;
; QY 165 HFLFKTSSGSTPLFSSSSPGYPLTSGTYVTPPRLPRNTFSKAFKLLKPKSKYCSWKCA 224
; DB 17 HELFK-PGTSPLFCTSPGYPLTSGTYVTPPRLPRNTFSKAFKLLKPKSKYCSWKCA 75
;
; QY 225 ALSATAAALLLAILLAYFI-----VPWSL----- 248
; DB 76 ALSAIVISATLVILLAYFVAMHLFGLNHLQPMEGOMYEITDASSWPVPTDVSILPSG 135
;
; QY 249 -----KNSSIDSGAEVGRVTVQVPPGVFWRSQIHISQPO 284
; DB 136 GTGLETPDRKGKGTGKPPSSFPEDSFIDSGEIDVGRRASQKIPPGTFWRQVFDHPV 195
;
; QY 285 FLKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK-----EKWSVYESPRRSIQT 339
; DB 196 HLKFNYSLGAALVGIYGRKGLPPSHQAQYDFMERLDGK-----EKWSVYESPRRSIQT 339
;
; QY 340 LVQNEAVFYQYLDVGLWHLAFYNDGKEMVSEFNTVVL 377
; DB 256 PSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAT 293
```

RESULT 7

```
US-08-491-357-3
; Sequence 3, Application US/08491357
; Patent No. 5716782
;
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,357
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
```

```
; ANTI-SENSE: NO
US-08-491-357-3

Query Match
Best Local Similarity 6.0%; Score 122; DB 1; Length 872;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;

QY 9 RKPAEAGRPPTPPSSPSSLLPSAQLPSS-HNP-PPVSCOMPLLDSTNTHQIMDTPDEEF 66
DB 68 KKPAARG--PGPPATPPO--POPSPQGVHTFPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPOQASSGPPNHHSSQSTLPPPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NYLVPTPSKTOOGLYQAFGNPQFQS-----PPAKQTSFYSKQTPHHS----- 158
QY 122 SLTNRROIHAPAPNDLATTPESVQLQDSWLNNSVPLETRHFLFKTSSGSTPLFSSS 181
DB 159 -----PPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPPRLPRNTFSRKAFKLLKPKSKYCSWKCAALSAIA 230
DB 172 GPGSPAQDIYQVPPSAGTGHDYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLLAILLAYFIVPWSLKNSSI-DSGEAE-----VGRVTOEYVP--GVF-- 273
DB 212 -----AKVVVTRVCGQYVVEASQAQDEYDTPRHLAPGSQDIYDVPVRGLLPN 262
QY 274 -WRSOIHSQFQFLKFNISLGDALFGY-----IRRLGPPS--HAQYD 314
DB 263 QYGOEYDTPDMAVKG--PNGRDPLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 8
US-08-968-633-3
; Sequence 3, Application US/08968633
; Patent No. 6100384
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/968,633
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4044
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: NO. 6100384 Relevant
; TOPOLOGY: NO. 6100384 Relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-968-633-3

Query Match
Best Local Similarity 6.0%; Score 122; DB 3; Length 872;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;

QY 9 RKPAEAGRPPTPPSSPSSLLPSAQLPSS-HNP-PPVSCOMPLLDSTNTHQIMDTPDEEF 66
DB 68 KKPAARG--PGPPATPPO--POPSPQGVHTFPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPOQASSGPPNHHSSQSTLPPPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NYLVPTPSKTOOGLYQAFGNPQFQS-----PPAKQTSFYSKQTPHHS----- 158
QY 122 SLTNRROIHAPAPNDLATTPESVQLQDSWLNNSVPLETRHFLFKTSSGSTPLFSSS 181
DB 159 -----PPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPPRLPRNTFSRKAFKLLKPKSKYCSWKCAALSAIA 230
DB 172 GPGSPAQDIYQVPPSAGTGHDYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLLAILLAYFIVPWSLKNSSI-DSGEAE-----VGRVTOEYVP--GVF-- 273
DB 212 -----AKVVVTRVCGQYVVEASQAQDEYDTPRHLAPGSQDIYDVPVRGLLPN 262
QY 274 -WRSOIHSQFQFLKFNISLGDALFGY-----IRRLGPPS--HAQYD 314
DB 263 QYGOEYDTPDMAVKG--PNGRDPLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 9
US-09-196-466-3
; Sequence 3, Application US/09196466
; Patent No. 6124434
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,466
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
```

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-196-466-3

Query Match 6.0%; Score 122; DB 3; Length 872;
Best Local Similarity 22.6%; Pred. No. 0.015;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;
QY 9 RKPASAGRIPIPTSPSLLPSAQLPSS-HNP-PPVSCQMPLLDSNTSHQIMDTNDEEF 66
DB 68 KKPAPG--PGPATPPQ--PQSLPGQVHTPPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQQAASSGPPNHHQSQTLLRPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NVILVPTPSKTQOGLYQAPGNPQFS-----PPAKQTSFQTPHHS----- 158
QY 122 SLTNRRSQIHAPAPANDLATTPEVSQLODSWVLSNVPLETRHFLFKTSSGSTPLFSS 181
DB 159 -----FPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSQT---VYTPPRLPRNTFSRKAFKLPKPSKYCSWKCAALSAIA 230
DB 172 GFGSPAQDIYQVPPSAGTGHDIYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLAILLAYFTVPWSLKNSSI-DSGEAE-----VGRVTOEVP--GVF-- 273
DB 212 -----AKVVPTRVGGYVYEASQAQDEYDTPRHLLAPGSQDIYDVPVRCGLLPN 262
QY 274 -WRSQIHISQPFLENISLGDALFGVY-----IRGLPPS--HAQYD 314
DB 263 QYQGEYDTPPMAVKG--PNGRDPDLLDYDVPSPVEKGLPPSNHHHSYVD 309

RESULT 10

PCT-US96-10823-3
Sequence 3, Application PC/TUS9610823
GENERAL INFORMATION:
APPLICANT: Golemis, Erica A.
APPLICANT: Law, Susan
APPLICANT: Estojak, JoAnne
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
TITLE OF INVENTION: ALTERATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10823
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US96-10823-3

Query Match 6.0%; Score 122; DB 5; Length 872;
Best Local Similarity 22.6%; Pred. No. 0.015;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;
QY 9 RKPASAGRIPIPTSPSLLPSAQLPSS-HNP-PPVSCQMPLLDSNTSHQIMDTNDEEF 66
DB 68 KKPAPG--PGPATPPQ--PQSLPGQVHTPPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQQAASSGPPNHHQSQTLLRPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NVILVPTPSKTQOGLYQAPGNPQFS-----PPAKQTSFQTPHHS----- 158
QY 122 SLTNRRSQIHAPAPANDLATTPEVSQLODSWVLSNVPLETRHFLFKTSSGSTPLFSS 181
DB 159 -----FPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSQT---VYTPPRLPRNTFSRKAFKLPKPSKYCSWKCAALSAIA 230
DB 172 GFGSPAQDIYQVPPSAGTGHDIYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLAILLAYFTVPWSLKNSSI-DSGEAE-----VGRVTOEVP--GVF-- 273
DB 212 -----AKVVPTRVGGYVYEASQAQDEYDTPRHLLAPGSQDIYDVPVRCGLLPN 262
QY 274 -WRSQIHISQPFLENISLGDALFGVY-----IRGLPPS--HAQYD 314
DB 263 QYQGEYDTPPMAVKG--PNGRDPDLLDYDVPSPVEKGLPPSNHHHSYVD 309

RESULT 11

US-08-894-454-163
Sequence 163, Application US/08894454
Patent No. 6544784
GENERAL INFORMATION:
APPLICANT: VAN DEN VEN, W.J.M.
APPLICANT: SCHOENMAKERS, H.F.P.M.
TITLE OF INVENTION: MULTIPLE-TUMOR ABERRANT GROWTH
TITLE OF INVENTION: GENES
NUMBER OF SEQUENCES: 164
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Webb Law Firm
STREET: 700 Koppers Building, 436 Seventh Avenue
CITY: Pittsburgh
STATE: PA
COUNTRY: USA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,454
FILING DATE: 15-AUG-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP/00716
FILING DATE: 19-FEB-1996
APPLICATION NUMBER: 95200390.3
FILING DATE: 17-FEB-1995
APPLICATION NUMBER: 95201951.1
FILING DATE: 14-JUL-1995

ATTORNEY/AGENT INFORMATION:
NAME: Johnson, Barbara E
REGISTRATION NUMBER: 31,198
REFERENCE/DOCKET NUMBER: 702-971100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 412-471-8815
TELEFAX: 412-471-4094
TELEX:
INFORMATION FOR SEQ ID NO: 163:
SEQUENCE CHARACTERISTICS:
LENGTH: 612 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-894-454-163

Query Match 5.8%; Score 117; DB 4; Length 612;
Best Local Similarity 24.6%; Pred. No. 0.026;
Matches 61; Conservative 20; Mismatches 95; Indels 72; Gaps 13;

QY 6 NGRKPSAAG---RPPTSPSLLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDTPN 62
DB 56 NPYKQPGEGDFLPPPPPLDSSALPSI---SGNFPPP-----PPLDEEAFK--VQGNP 105
QY 63 -----DEEFPNSYLLR--ACSGPQQASSSGPPNNHHSQSTLRPLPPPHNHTL 108
DB 106 GKGTEERSSLDADISLTSLADLECSPYKPR---PQSGTGSTASPPVTPVT--- 159
QY 109 SHHSSANSLNRRSLNRRSQIHAPAPAP-----NDLATTPEVSQVQLDQSWLNSN 158
DB 160 --GKRWVNPQPLTATKSTLKPAPAPQAGPIVAPIGTLKQPQVPVASYTTASTSS 217
QY 159 VPL-----ETRHLEFKTSSGTPLFSSSSPGY-----PLTSGTV---Y 193
DB 218 RPTENVVKSQSPHYMAAPSSQ--IYSGPQGYNTQPPVPSGQCQPPPTRGMDVAY 275
QY 194 TPPRLLP 201
DB 276 IPPGQLP 283

RESULT 12
US-08-264-534-34
Sequence 34, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,534
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-264-534-34

Query Match 5.8%; Score 117; DB 1; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;

QY 7 KGRKPSAAGRPPTSPSLLPSAQ-LPSSHNPVPSVSCQMPLLDSNTSHQIMDTPDEE 65
DB 288 KSRPSAKS---TWPTSLPNAKEAKDAKGRRKKSLSERKVSQSVTLSPVDSLE--- 341
QY 66 FSPNSY-----LLRACSGPQQASSSGPPNNHHSQSTL-----RPLP--- 101
DB 342 -SPHTYVSDTTSSPMITSPGILQASPNPLATAAPPAPVHAQHALSFNLSHMQPLANGA 400
QY 102 ----PPHNHTLSHH-----SSANSLNRRSLNRRSQIHAPAPAPNDLATTPEVSQVQLD 151
DB 401 STVLPSVSQLSHHHIIVSPGSGSAGLSR-----LH-PVPVPAD----- 438
QY 152 SWLNSNVPLTRH-----FLFKTSSGSTPLFSSSPGYPLTSGTVTP---PB---R 198
DB 439 -WMRMENV-ETQNMFGMWLAPEAGTHPGIAPOS--RPPEGKHITTPRPLPIVTFQ 494
QY 199 LLPRNTESRKAFKLKPKSKYCSWKCAALSAIAAALLAILLAYFTVPSLKNSSIDSSEA 258
DB 495 LIPKGSIAQA-GAPQPOSTCP-----PAGVP-----LTMQIIP----- 529
QY 259 EVGRVTVQEVPPGVFWRSQIHISO---PQFLKFNISLQKDALFGVYIRRGLPSPSHAQY 313
DB 530 EMARLPSVAFPTAMMPQDQGVQVQOTILPAYHPFASVKG-----YPTPPSQHSY 578

RESULT 13
US-08-083-590A-13
Sequence 13, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,590A
FILING DATE: 25-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090

TELEFAX: 212 869864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-083-590A-13

Query Match 5.8%; Score 117; DB 1; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;
QY 7 KGRKPSAAGRPPTSPSLPSAQ-LPSSHNPVPVSCOMPLDSDNTSHQIMDTPDEE 65
DB 288 KSRPSSAKS---TMPSTLNLAKAKDAKSRKKSLSKVKQLSESVTLSPVDSLE--- 341
QY 66 FSPNSY-----LLRACGPOQASSGPPNHHQSSTL-----RPPLP--- 101
DB 342 -SPHTVSDTSSPMITSPGILQASPNMLATAAPVPAVHAQHALSFNLHEMQLAHGA 400
QY 102 ----PPHNTLSHH-----SSANSLNRSLNRSQIHAPAPNDLATTTPESVOLQD 151
DB 401 STVLPVSQOLLHHHIVSPGSGSAGLSR-----LH-PVPVPAD----- 438
QY 152 SWVLNSNVPLETRH-----FLFKTSSGSTPLFSSSPGYPLTSGTYTP---PP---R 198
DB 439 -WMNRMEVN-ETQYNEMFQWLAPAEHTPGIAPOS--RPPEKHHITTPREPLPIVTFQ 494
QY 199 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
DB 495 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
QY 259 EYGRRTVQVPPGVFWRSGIHSQ---POFLKFNISLGDALFGVYIRRGLPSPSHAY 313
DB 530 EMARLPSVAFPTAMPPODQOVAQTILPAYHPPASVVGK-----YTPPSQHSY 578

RESULT 14
US-08-465-500-34
Sequence 34, Application US/08465500
Patent No. 5789195
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Muskavitch, Marc A.T.
APPLICANT: Fehon, Richard G.
APPLICANT: Rebay, Ilaria
APPLICANT: Blauweller, Cristine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,500
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-034

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-465-500-34

Query Match 5.8%; Score 117; DB 1; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;
QY 7 KGRKPSAAGRPPTSPSLPSAQ-LPSSHNPVPVSCOMPLDSDNTSHQIMDTPDEE 65
DB 288 KSRPSSAKS---TMPSTLNLAKAKDAKSRKKSLSKVKQLSESVTLSPVDSLE--- 341
QY 66 FSPNSY-----LLRACGPOQASSGPPNHHQSSTL-----RPPLP--- 101
DB 342 -SPHTVSDTSSPMITSPGILQASPNMLATAAPVPAVHAQHALSFNLHEMQLAHGA 400
QY 102 ----PPHNTLSHH-----SSANSLNRSLNRSQIHAPAPNDLATTTPESVOLQD 151
DB 401 STVLPVSQOLLHHHIVSPGSGSAGLSR-----LH-PVPVPAD----- 438
QY 152 SWVLNSNVPLETRH-----FLFKTSSGSTPLFSSSPGYPLTSGTYTP---PP---R 198
DB 439 -WMNRMEVN-ETQYNEMFQWLAPAEHTPGIAPOS--RPPEKHHITTPREPLPIVTFQ 494
QY 199 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
DB 495 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
QY 259 EYGRRTVQVPPGVFWRSGIHSQ---POFLKFNISLGDALFGVYIRRGLPSPSHAY 313
DB 530 EMARLPSVAFPTAMPPODQOVAQTILPAYHPPASVVGK-----YTPPSQHSY 578

RESULT 15
US-08-346-128-34
Sequence 34, Application US/08346128
Patent No. 5856441
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,128
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,038
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-346-128-34

Query Match 5.88; Score 117; DB 2; Length 657;
Best Local Similarity 22.18; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;

QY	7	KGRKPSAAGRPITPTSSPILLSAQ-LPSSHPPVSCOMPLDSDNTSHQIMDTPDEE	65
Db	288	KSRPSAKS---TMPTSLNLAKAKDAKGRKKSLSKVKQLSESVTLSPVDSLE---	341
QY	66	FSPNSY-----LLRACSGPQQAASSGPPNHHQSSTL-----RPPLP---	101
Db	342	-SPHTVSDTSSPMITSPGILQASPNPLATAAPAPVHAQHLSFSLNHEMQLAHGA	400
QY	102	-----PPHNTLSHHH-----SSANLNNSLTNRRSQIHAPAPAPNDLATTPESVQLOD	151
Db	401	STVLPVSQQLSHHHITVSPGSGSAGLSR-----LH-PVPVPAD-----	438
QY	152	SMVLNSNVPLETRH-----FLKTSSTGTPLFSSSPGYPLTSGTVYTP-----PP----	198
Db	439	-WNRMEVN-ETQYNEMFCWVLAPAEHTHPGIAPOS--RPPEGKHITTPRELPPPIVTFQ	494
QY	199	LIPRNTFSRKAFKPKSKYCSWKCAALSAIAAAILLAILLAYFIVPWSLKNSSIDSGEA	258
Db	495	LIPKGSIAQPA-GAPQPSTCP-----PAVAGP-----LPTMYQIP-----	529
QY	259	EVGRRVTQEPVPGVFRSQIHISQ----POFLKFNISLGKDALFGVYIRGLPPSHAQY	313
Db	530	EMARLPSVAFPTAMMPQODGQVAQTILPAYHPFPASVGR-----YPTPPSQHSY	578

Search completed: August 26, 2003, 14:32:51

Job time : 23 secs

THIS PAGE BLANK (user)